AFRL-IF-RS-TR-2004-5
In-House Technical Report
January 2004

RESTORATION OF WAVELET-COMPRESSED IMAGES AND MOTION IMAGERY

Mark A. Robertson

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2004-5 has been reviewed and is approved for publication.

APPROVED: /s/
GERALD C. NETHERCOTT, Chief
Multi-Sensor Exploitation Branch
Information & Intelligence Exploitation Division

FOR THE DIRECTOR: /s/
JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	JANUARY 2004	**In-House, January 2003 – August 2003**

4. TITLE AND SUBTITLE

Restoration of Wavelet-Compressed Images and Motion Imagery

5. FUNDING NUMBERS

C - N/A
PE - G32605
PR - 558B
TA - CI
WU - TE

6. AUTHOR(S)

Mark A. Robertson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

AFRL/IFEC
32 Brooks Road
Rome, NY 13441-4114

8. PERFORMING ORGANIZATION REPORT NUMBER

AFRL-IF-RS-TR-2004-5

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL (CITE)
c/o Bruce Suter (AFRL/IFGC)
525 Brooks Road
Rome, NY 13441-4114

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2004-5

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Mark A. Robertson/IFEC/315-330-4171/Mark.Robertson@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

This technical report investigates the characteristics of compression noise in images and motion imagery compressed by scalar quantization of the data's two- or three-dimensional wavelet transform coefficients. Such quantization noise is both experimentally and theoretically shown to be spatially varying in the pixel domain, with statistical correlations between the errors at the pixel locations. A quantization noise covariance matrix is presented that can find use in general restoration scenarios where the observed image or images have been compressed by scalar quantization of the data's wavelet coefficients. Several restoration examples, including de-blurring for the single-image case and temporal filtering for the motion-imagery case, are provided to demonstrate the quantization noise model's advantage over the common assumption of independent and identically distributed noise.

14. SUBJECT TERMS

quantization noise, compression error, image and video compression, discrete wavelet transform, deblurring, temporal filtering, JPEG 2000, SPIHT

15. NUMBER OF PAGES

68

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Contents

List of Figures

List of Tables

Acknowledgments

The author would like to acknowledge the support of the Center for Integrated Transmission and Exploitation (CITE), a joint endeavor of the Information Directorate and the Air Force Office of Scientific Research, both of the Air Force Research Laboratory. The author would also like to thank Dr. Andrew Noga for helpful comments in preparation of this manuscript.

Summary

Most imagery and motion imagery data are compressed in a lossy fashion, implying that the available data are not exact replicas of the data observed by the sensor prior to compression. While such compression errors are not typically noticed by the viewer, there are numerous image and video processing algorithms that are very sensitive to noise. Compression errors have very noise-like characteristics, and if one does not properly account for them the algorithms may fail to give satisfactory results. Algorithms that are highly sensitive to noise typically require extremely accurate models of how the data were acquired, and lossy compression forms an integral component of the acquisition process whose influence must not be ignored.

The current document examines this compression noise for the situation when the compression makes use of scalar quantization of the data's wavelet coefficients. Application scenarios include, but are not limited to, the following: imagery compressed according to the international JPEG 2000 standard; imagery compressed according to the SPIHT or EZW compression techniques; motion imagery compressed according to the international Motion JPEG 2000 standard; and motion imagery compressed according to the three-dimensional SPIHT technique.

Compression errors have unique behavior. This technical report provides a statistical characterization of the compression noise which can be used in general scenarios that are sensitive to noise in the data. Two example applications are presented: image deblurring for the case of a single image, and temporal filtering for the case of motion imagery. Results suggest that improvements for these applications are obtainable by using the statistical characterizations relative to simpler models; the cost of such improvements is the additional complexity of the more accurate model. Other situations that may benefit from this work include the following: super-resolution image reconstruction; deconvolution; image segmentation and classification; and statistical pattern and target recognition.

1. Introduction

Using wavelet transforms for compression of images is becoming increasingly popular. A two-dimensional discrete wavelet transform (DWT) is often applied for compression of still images, as well as for individual frames of a motion imagery sequence. Three-dimensional wavelet transforms are also applied to imagery volumes such as motion imagery sequences and hyperspectral imagery data. In either case, when such compression is lossy there is an error in the reconstructed image due to quantization of the wavelet coefficients. In many image processing applications, it may be beneficial to be able to characterize the reconstruction error—for example, numerous algorithms in image processing use probabilistic models for the observation noise, and more accurate probabilistic models lead to more accurate results from the overall algorithm. Characterization of compression error is especially important when such error is the dominant source of noise in a system.

Compression noise has been studied thoroughly for the case of the discrete cosine transform (DCT) [1], which found and continues to find extensive use in compression of images and video. However, wavelet compression noise has not been studied in as much detail. Woods and Naveen [2] considered the compression distortion for purposes of bit allocation, but consider the error from a collective point of view (total error energy being the sum of the total error energy from each subband) rather than the local point of view (down to the pixel level) taken here. To see the limitations of such a global point of view, consider the image of Figure 1, which shows the mean-squared error averaged over 1320 images of size 256×256 that were compressed according to the wavelet-based JPEG 2000 standard [3]. Bright pixels in the figure correspond to higher mean-square errors at those pixel positions. An obvious conclusion from the image is that the variances of the errors are changing according to pixel location, i.e., the pixel errors are not identically distributed.

Not only are the compression errors reported in Figure 1 spatially varying, but they are correlated as well. Table 1 shows the correlation of the error at pixel $(66, 66)$ with the errors of surrounding pixels, which clearly demonstrates that there are significant correlations in the quantization error that should not be ignored by algorithms that process this compressed imagery. Although in practice it is tempting to assume independent and

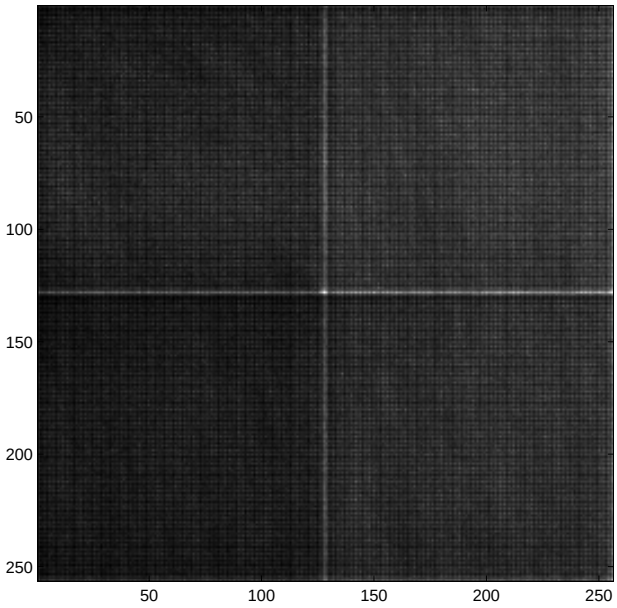


Fig. 1. Mean-squared error for 1320 images of size 256×256 compressed according to JPEG-2000 at 10:1 compression ratio with 128×128 tiles, with five levels of wavelet decomposition. The image has been slightly brightened to emphasize details. Note that the block box surrounding the image is part of the axis, and not part of the error image.

identically distributed (IID) noise due to its simplicity, the examples just shown, in addition to the derivations and experimental evidence to follow in later sections, demonstrate that the quantization error is not independent nor identically distributed, and proper consideration of the noise's statistical characteristics can be beneficial.

Error patterns such as that shown in Figure 1 are readily evident when applying a three-dimensional wavelet transform to image volumes. However, in this case there are peaks and valleys in the mean-squared error in the two spatial dimensions as well as the third dimension. Examples for the three-dimensional case are given in later sections.

The tile boundary effect that is apparent in Figure 1 is well known, and there are various methods of counteracting such a problem [4][5]. Nevertheless, even if one carefully compresses an image so as not to produce excessive boundary effects, the error throughout a compressed tile will still exhibit spatial-domain errors that vary in a manner similar to that of Figure 1.

In general, the quantization error in the pixel domain will be the sum of the product of the wavelet coefficients' quantization errors times the corresponding basis image for each

TABLE 1

Error correlations for a single pixel location from the experiment in Figure 1. The peak center value corresponds to the mean-square error from the (66,66) location in Figure 1, while, for example, the entry below the peak corresponds to the correlation between errors at locations (66,66) and (67,66).

-1.23	-0.77	0.10	-2.71	-2.54	-1.54	-0.16	0.66	1.28
1.57	-0.72	0.40	-1.49	0.33	-0.15	0.24	0.21	0.81
-1.94	-1.53	0.98	0.35	1.18	0.73	0.11	-0.42	-0.47
-3.82	-0.84	-1.08	2.57	10.59	1.97	-1.59	-0.33	0.22
-2.59	0.04	-2.59	5.84	22.49	6.63	-1.58	-2.44	-1.14
-2.12	-0.82	-1.04	3.38	10.73	4.42	-1.53	-2.90	-2.55
-1.31	-1.85	-1.29	0.79	0.89	1.44	-0.67	-0.40	-1.24
1.08	1.60	1.87	-1.09	-2.79	1.15	0.38	0.95	1.51
1.37	0.09	1.76	-0.67	-3.55	0.78	1.33	1.37	0.56

coefficient. Such a statement is obvious and has been pointed out, for example, by Watson et al. [6]. However, the authors of [6] were primarily concerned with the perceptibility of wavelet quantization noise, whereas the focus of this paper is a probabilistic examination of the quantization noise. Such a probabilistic model of the quantization noise could then be used, for example, in imagery or video restoration.

The next section provides an analysis of the spatial-domain effect of wavelet quantization noise, the main result of which is a covariance matrix for the error that can be used in conjunction with various probability distribution functions (pdf). Section 3 discusses implications of Section 2, including example predictions of quantization noise behavior based on the noise model, as well as further analysis of the observed pixel errors reported in Figure 1. Section 4 provides an example restoration formulation for image deblurring that makes use of the quantization noise model. A further example of temporal filtering for the case of motion imagery is presented in Section 5. Results in both example cases demonstrate that using the proposed quantization error model allows improvements relative to the simple but common independent and identically distributed noise model. Finally, conclusions are presented in Section 6.

2. Quantization Noise

Without loss of generality, we represent the pixel-domain image data by the length-N vector $\mathbf{z}$, which is formed by stacking the columns from either a two-dimensional image or the images of a three-dimensional imagery volume. For notational convenience it will be assumed that $\mathbf{z}$ represents a single $W \times W$ image; generalizations to non-square images and motion imagery are straightforward. The multiresolution DWT being employed is represented by the matrix $\mathbf{H}$, of size $N \times N$, which can be either the two- or three-dimensional DWT, depending on the situation. The wavelet coefficients are given as $\mathbf{y} = \mathbf{Hz}$, which are quantized to $\mathbf{y}_q = \mathbf{Hz} + \mathbf{e_y}$, where $\mathbf{e_y}$ represents the error due to quantization of the wavelet coefficients. Upon application of the inverse wavelet transform, the reconstructed image becomes $\mathbf{z}_q = \mathbf{z} + \mathbf{e_z}$, where the spatial domain quantization error is $\mathbf{e_z} = \mathbf{H}^{-1}\mathbf{e_y}$, where $\mathbf{H}^{-1}$ is the inverse DWT.

An alternative but equivalent way of looking at an image is in terms of the basis images of the inverse DWT. Then,

$$\mathbf{z} = \sum_{u,v} \mathbf{h}_{u,v}^{-1} Y[u,v], \tag{1}$$

where $Y[u,v]$ is the $(u,v)^{th}$ element of the DWT decomposition represented by $\mathbf{y}$, and $\mathbf{h}_{u,v}^{-1}$ is the basis image corresponding to the $(u,v)^{th}$ wavelet coefficient, equivalent to the $(vW+u)^{th}$ column of $\mathbf{H}^{-1}$. In a similar manner, the quantization error can be written as

$$\mathbf{e_z} = \sum_{u,v} \mathbf{h}_{u,v}^{-1} E_Y[u,v]. \tag{2}$$

Watson et al. [6] base much of their quantization noise perceptibility study on the error terms $\mathbf{h}_{u,v}^{-1} E_Y[u,v]$.

While the quantization error is a deterministic function of an input image, in many practical applications once the quantized signal $\mathbf{y}_q$ is calculated, the clean signal $\mathbf{y}$ is discarded, and thus explicit information about the quantization error is lost. A commonly used theoretic tool for modeling the error signal is to treat it as a random quantity [7]. Treating the error as random provides an understanding of how the error behaves, including how much the error will vary at different pixel locations and the correlations between the errors. Such understanding further provides the theoretical framework for formulating effective schemes for alleviating the error. This statistically-based model of the error signal is referred to as

a quantization noise model since the error signal represents unwanted information in the resulting image representation. In the context of DCT-based image compression, there are numerous cases that treat the compression error as a random quantity; some do so in order to analyze visibility [8] or characteristics [9] of the error, while others do so in order to formulate algorithms that attempt to remove the noise [10][11][12]. This work is interested in both the characterization and the alleviation of the compression noise.

The original image data $\mathbf{z}$ has a covariance matrix of $\mathbf{K_z}$, which results in a covariance matrix for the wavelet coefficients of $\mathbf{K_y} = \mathbf{H}\mathbf{K_z}\mathbf{H}^t$. For images well-modeled by a separable extension of first-order stationary Markov sequences with positive correlation parameter ρ (a simple but common image model [13]), many wavelet transforms will approximately provide uncorrelated transform coefficients. The coefficients are not strictly uncorrelated, as is readily evident by looking at the lowest frequency band of a wavelet decomposition; the decorrelating effect also depends on the number of levels of the wavelet decomposition. Here, it is assumed that the coefficients are approximately uncorrelated, allowing the simplifying approximation that $\mathbf{K_y}$ is diagonal, which leads to the approximation that the covariance matrix of $\mathbf{e_y}$, $\mathbf{K_{e_y}}$, is approximately diagonal, and consists of the wavelet domain quantization error variances for each coefficient. Given $\mathbf{K_{e_y}}$, the covariance of the quantization error in the spatial domain can then be found as

$$\begin{aligned} \mathbf{K_{e_z}} &= \mathbf{H}^{-1}\mathbf{K_{e_y}}\mathbf{H}^{-t} \\ &= \mathbf{H}^{-1}\left[\mathbf{H}^{-1}\mathbf{K_{e_y}}\right]^t, \end{aligned} \tag{3}$$

showing that the error covariance in the pixel domain depends only on the error variances in the wavelet domain and the basis images of the inverse wavelet transform. Individual elements of $\mathbf{K_{e_z}}$ can be easily expressed in terms of the basis images $\mathbf{h}_{u,v}^{-1}$,

$$K_{\mathbf{e_z}}[m_1, m_2] = \sum_{u,v} \sigma_Y^2[u,v] h_{u,v}^{-1}[m_1] h_{u,v}^{-1}[m_2], \tag{4}$$

where $\sigma_Y^2[u,v]$ are the DWT-domain error variances that compose $\mathbf{K_{e_y}}$, and $h_{u,v}^{-1}[m]$ represents the m^{th} element of $\mathbf{h}_{u,v}^{-1}$. A special case of (4) considers the diagonal elements of $\mathbf{K_{e_z}}$, which represent the variances of the pixel-domain quantization errors. Quantization error variances are themselves of significant interest; the results of Figure 1 represent estimates

of these terms for one set of data. In general,

$$\mathrm{diag}(\mathbf{K}_{\mathbf{e_z}}) = \sum_{u,v} \sigma_Y^2[u,v] \left[\mathbf{h}_{u,v}^{-1}\right]^2, \tag{5}$$

where the square of the vector indicates that each element of the vector is squared, and $\mathrm{diag}(\mathbf{K}_{\mathbf{e_z}})$ is the vector taken from the diagonal of $\mathbf{K}_{\mathbf{e_z}}$. The $\left[\mathbf{h}_{u,v}^{-1}\right]^2$ terms above can be considered "error variance basis images," since the error variances of an image can be written in terms of the basis summation of (5). Examples of such variance basis images are given in the next section.

It is argued here that a Gaussian probability distribution function provides a good description of the quantization error in the pixel domain. The primary justification is due to the basis image summation that forms a reconstructed pixel error—the quantization error for a single pixel will consist of the sum of quantization errors for each basis image that overlaps with the pixel, as described by (2). The number of such basis images depends on several factors, including the length of the wavelet reconstruction filters, the levels and type of wavelet decomposition, and the location of the pixel within the fixed-length transform size. For most situations of interest, the number of contributing terms is large enough such that one may use the Central Limit Theorem to approximate the sum of random variables represented by noisy basis images as Gaussian. Thus, the probability distribution function of the pixel-domain quantization error is approximated as

$$p(\mathbf{e_z}) = \frac{1}{(2\pi)^{N/2} \left|\mathbf{K}_{\mathbf{e_z}}\right|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{e}_\mathbf{z}^t \mathbf{K}_{\mathbf{e_z}}^{-1} \mathbf{e_z}\right\}. \tag{6}$$

The exponent in the above equation can be expanded according to (3),

$$-\frac{1}{2}\mathbf{e}_\mathbf{z}^t \mathbf{K}_{\mathbf{e_z}}^{-1} \mathbf{e_z} = -\frac{1}{2}\left(\mathbf{H}\mathbf{e_z}\right)^t \mathbf{K}_{\mathbf{e_y}}^{-1} \left(\mathbf{H}\mathbf{e_z}\right), \tag{7}$$

showing that the exponent of the probability distribution can be evaluated by simple application of the wavelet transform followed by scaling according to the diagonal $\mathbf{K}_{\mathbf{e_y}}^{-1}$.

Other distributions besides Gaussian are possible, although the Gaussian distribution is certainly the most convenient both from a computational standpoint and from the standpoint of including arbitrary covariance matrices as is necessary in the problem at hand. One alternative to a Gaussian distribution is a multivariate Laplace distribution, which more

closely models heavy tails of a noise source. Various definitions of a multivariate Laplace distribution exist; Kotz et al. [14] define it in terms of its characteristic function $\mathbf{\Phi}(\omega)$,

$$\mathbf{\Phi}(\omega) = \frac{1}{1 + \frac{1}{2}\omega^t \mathbf{K}_{\mathbf{e_z}} \omega}. \tag{8}$$

Unfortunately, using (8) in a general image or video restoration formulation can be problematic due to the probability density function corresponding to the above characteristic function—the explicit density involves Bessel functions, which can prove difficult from a computational point of view in typical maximum likelihood (ML) or maximum a posteriori (MAP) restoration formulations.

From a computational standpoint, a more reasonable version of the multivariate Laplace distribution is given by the following, taken from [15]:

$$p(\mathbf{e_z}) = \frac{K_{N,\eta}}{|\mathbf{C}|^{1/2}} \exp\left\{-[\mathbf{e_z}^t \mathbf{C}^{-1} \mathbf{e_z}]^{\eta/2}\right\}, \tag{9}$$

where $K_{N,\eta}$ is a constant depending on η and N [15]. Equation (9) is a general distribution parameterized by η; choosing $\eta = 2$ yields the common multivariate Gaussian distribution, while choosing $\eta = 1$ yields a multivariate Laplace distribution. For the situation at hand we are interested in $\eta = 1$, for which case the covariance of $\mathbf{e_z}$ in (9) is $(N+1)\mathbf{C}$. Choosing $\mathbf{C} = \mathbf{K}_{\mathbf{e_z}}/(N+1)$ thus yields the desired covariance for the error $\mathbf{e_z}$ in (9).

Although not recommended here, it is possible to use independently-distributed Gaussian or Laplacian noise sources to model the quantization noise in the spatial domain. The obvious disadvantage of such a method is that the correlation structure contained in $\mathbf{K}_{\mathbf{e_z}}$ is lost. However, by assigning the variances of these N independent noise terms according to the elements of the diagonal elements of $\mathbf{K}_{\mathbf{e_z}}$ as in (5), one can still capture the variance behavior that was illustrated in Figure 1. The main advantage of such an approach is a potential reduction in computational complexity.

Given a distribution type, the individual wavelet domain quantization error variances that compose $\mathbf{K}_{\mathbf{e_y}}$ determine the error in the pixel domain. For high-rate situations, the quantization step sizes used for compression are small enough that a uniformly distributed random variable accurately models the quantization error in the wavelet domain; Watson et al. [6] also use this model. In such cases, the diagonal components of $\mathbf{K}_{\mathbf{e_y}}$ are composed

of terms $\frac{1}{12}\Delta_i^2$, where Δ_i is the quantization step size for the i^{th} wavelet coefficient. For lower-rate situations, one must resort to more complicated means: A distribution function must be assumed for each wavelet coefficient, whose parameters must be estimated from the received (noisy) data; the quantization noise can then be calculated for each wavelet coefficient. See [1] for such an analysis in the case of DCT quantization noise, where a Laplacian distribution is assumed for each DCT coefficient. Although a lower-rate analysis for the case of the wavelet transform could be performed analogously to that of the DCT, such situations were not the focus of this work, and are not considered further. In general, however, the original image statistics dictate the signal energy of the wavelet coefficients, and hence the energy of the quantization errors in the wavelet domain; such analysis is an interesting area for further investigation.

3. Discussion

Figure 2 shows the predicted PSNR for each of 16 video frames that are compressed by uniformly quantizing the coefficients from a temporal three-level wavelet decomposition with the 9/7 biorthogonal wavelet filters (note that this example only considers the overall frame-by-frame error, not the pixel-by-pixel error). The plot correlates very well with PSNR results reported for the 3D SPIHT video compression algorithm [16], which compresses imagery volumes by quantizing their 3D DWT coefficients. Such an analysis of frame-by-frame quantization error was previously performed by Xu et al. [17], who propose a wavelet transform that eliminates the low PSNR's at the group-of-frames boundaries that are evident in Figure 2. Nevertheless, regardless of whether or not a wavelet transform exhibits boundary effects, there will still be a predicted error pattern that depends on the reconstruction transform's basis vectors. It is the variability of error from frame to frame that can be quite important for an algorithm that is processing the data.

Figure 3 shows error histograms from the experiment reported in Figure 1. Two histograms are shown—one for the $(0, 0)$ location of every 4×4 block in the images, and the other for the $(1, 1)$ location of every 4×4 block. If one treats these histograms as probability mass functions, the variance for the $(1, 1)$ location is approximately twice that of the $(0, 0)$ location,

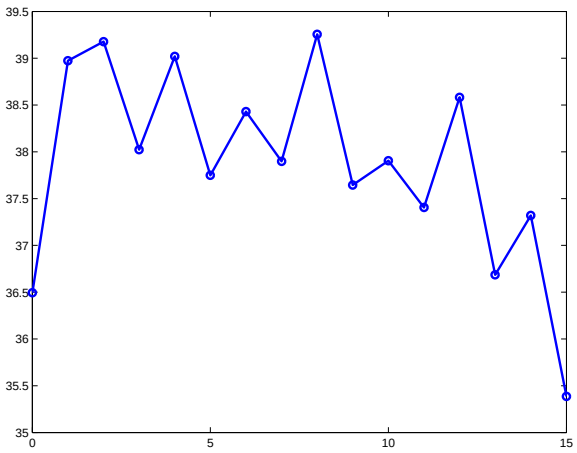


Fig. 2. Predicted PSNR for 16 frames when the wavelet coefficients of a temporal three-level 9/7 biorthogonal wavelet decomposition are quantized with uniform quantizers at high rate.

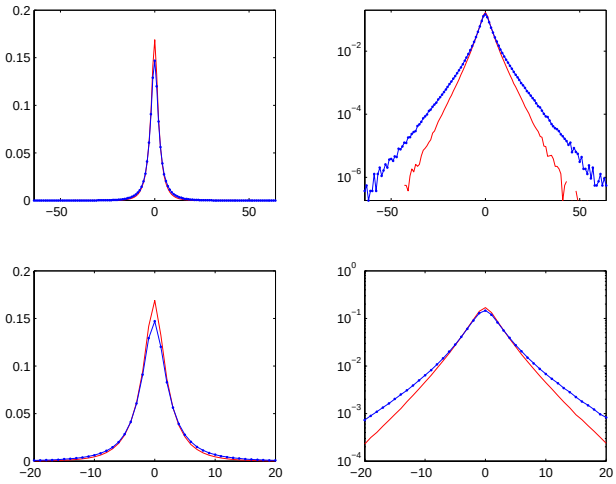


Fig. 3. Normalized error histograms for the test sequence in Figure 1. The solid line represents errors for pixels at locations (x, y) such that $x \bmod 4 = 0$ and $y \bmod 4 = 0$; the connected-dot line represents errors for pixel locations (x, y) where $x \bmod 4 = 1$ and $y \bmod 4 = 1$. The left-most plots show the error histograms at two different horizontal magnifications, and the right-most plots show the left-most plots on a logarithmic vertical axis.

again suggesting that not all pixel errors behave in the same manner. From the logarithmic histogram plots, it is apparent that these error distributions are not closely following a Gaussian; if they were, they would appear as inverted parabolas. One might argue that they more closely follow a Laplacian or Generalized Gaussian distribution, although the convexity changes that are apparent in the log-scale plots suggests otherwise. Although the plots of Figure 3 might tempt one to discard the Gaussian quantization noise model in favor of a heavier-tailed distribution, such an action is sometimes unnecessary—as discussed in the next sections, for several cases of restoring wavelet-compressed images the Gaussian noise model outperforms other Laplacian noise models. Note also that these tests are for one particular compression ratio and one particular data set, and the shapes of the distributions may vary at different compression ratios.

Figure 4 shows predicted quantization error variances for a situation similar to that shown in Figure 1, where the prediction is taken from the diagonal components of $\mathbf{K}_{\mathbf{e_z}}$. Uniform dead-zone scalar quantization is applied to each wavelet coefficient of subband b

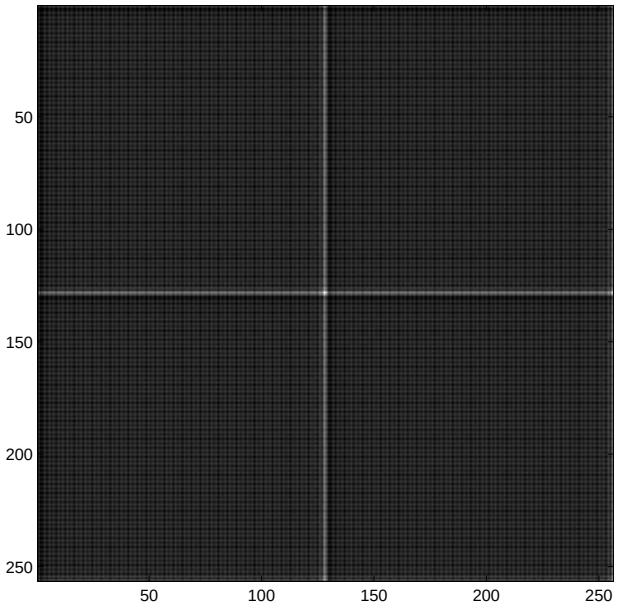


Fig. 4. Predicted error variance for compression of four tiles, taken from diag($\mathbf{K}_{\mathbf{e_z}}$).

with quantization step sizes

$$\Delta_b = \frac{\Delta_d}{\sqrt{\gamma_b}}, \tag{10}$$

where γ_b is the sum of the squared errors introduced by a unit error in subband b at a location away from the boundaries, and scales the nominal quantizer Δ_d [2]. The Daubechies 9/7 DWT [18] is the transform employed. A comparison between zoomed portions of Figures 1 and 4 is presented in Figure 5; a remarkable similarity between observed and predicted error variances is apparent. Note that the predictions shown in Figures 4 and 5 are based on the assumption that all wavelet coefficients are non-zero, and thus the larger quantization bins for zero-valued coefficients (due to the dead zone) do not contribute.

While Figure 2 showed only the temporal variation in error for compression using a three-dimensional DWT, there is also significant pixel variation as was present in Figure 4. Figure 6 shows both spatial and temporal mean-square error prediction for a 64×64 sequence of 16 frames, which has been compressed according to quantization of a three-level (both spatial and temporal) 3D wavelet decomposition. The sixteen individual error frames have been displayed in a 4×4 tile, with the first image at the top-left and the last image at the

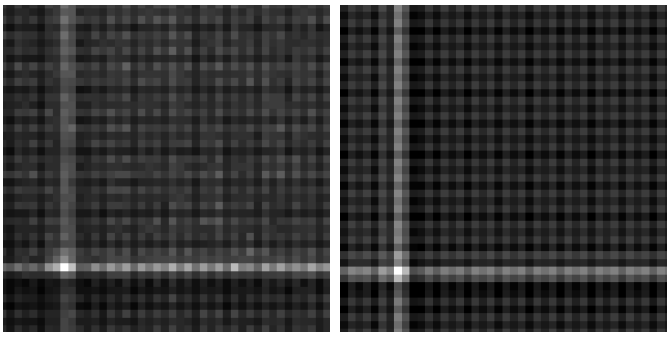


Fig. 5. Zoomed portions of Figures 1 (left) and 4 (right), near the centers of the images. (The ranges of both images have been normalized to the same range for display purposes.)

bottom-right, with intermediate images occurring in row order between these two. The overall brightness of each individual tile directly corresponds to the errors depicted in Figure 2.[1] However, Figure 6 clearly demonstrates that there is significant variation in the errors according to both temporal and spatial pixel location.

The variance basis images determine how the overall pixel-domain error variances will behave, as described by (5). Figure 7 shows four such basis images, taken from each of the four subbands of the lowest level of a five-level wavelet decomposition for 128×128 images. Summations of variance basis images such as these lead to overall variance images such as the one shown in Figure 4. Basis images at higher-frequency decomposition levels appear similar to those shown in the figure, but with less spatial extent. Since the example in the figure is for 128×128 images, there are 128^2 different basis images in total; however, for each subband most of these basis images are simple translates of each other.

Table 2 shows two predictions of the covariance of the error for the $(66, 66)$ position, as was reported for experimental observations in Table 1. The predictions have been normalized such that the variance value is 100.0. Again, the peak center value corresponds to the variance of the error from the (66,66) location. For part (a) of the table, predictions are computed assuming no zero-valued coefficients were received by the decoder, and hence no

[1]Note that Figure 2 plotted PSNR, where higher values indicate higher quality, whereas Figure 6 plots mean-square error, where higher values indicate poorer quality.

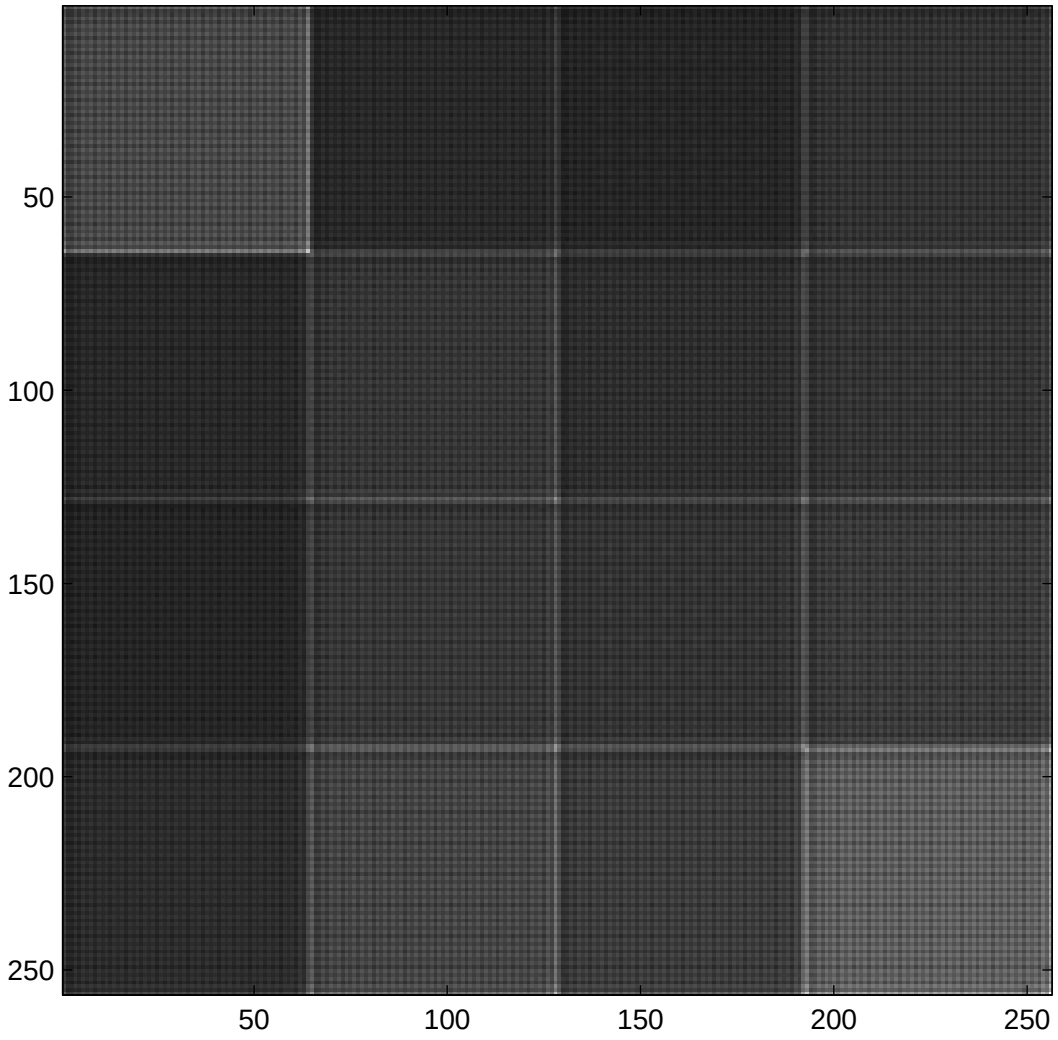


Fig. 6. Predicted quantization error variance for a length-16 volume of 64×64 images, compressed using a three-level (both temporal and spatial) wavelet decomposition. The 64×64 error variance images are tiled from top-left to bottom-right.

larger quantization bins due to the quantizer dead zone. In part (b) of the table, predictions are computed by randomly assigning zero and non-zero coefficients, and hence the larger quantization bins for dead-zone coefficients affect the result (note that this prediction is just one realization of many possible zero/non-zero coefficient scenarios, and different correlations occur for different configurations of zero/non-zero observations). Predicted correlations for part (a) of the table indicate only minor correlations between the errors at pixel $(66, 66)$

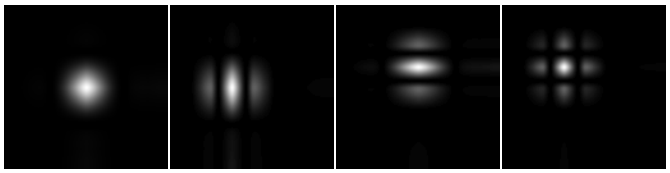


Fig. 7. Four error variance basis images, taken from the four subbands of the lowest level of a five-level wavelet decomposition for 128×128 images (black represents 0).

and its neighbors, and one might argue that approximating them as independent would be valid. However, as part (b) of the table clearly demonstrates, different quantization bin sizes that arise as a result of dead-zone quantization can introduce significant correlations between errors, and for this particular example such correlations exist beyond the eight nearest neighbors.

TABLE 2

Normalized prediction of error correlations for a single pixel location under the same conditions reported in Figure 1. (a) Predictions are computed assuming no zero-valued quantized coefficients; (b) predictions are computed by randomly assigning zero and non-zero quantized coefficients.

-0.27	-0.15	-0.62	-3.03	-6.36	-2.81	-0.21	0.30	0.18
-0.15	-0.38	-1.09	-2.51	-5.68	-2.15	-0.43	0.36	0.58
-0.62	-1.09	-0.95	-0.63	4.91	-0.14	-0.06	-0.10	0.35
-3.03	-2.51	-0.63	2.81	10.16	3.35	0.34	-1.42	-1.96
-6.36	-5.68	4.91	10.16	100.00	10.72	5.92	-4.56	-5.27
-2.81	-2.15	-0.14	3.35	10.72	3.90	0.85	-1.03	-1.73
-0.21	-0.43	-0.06	0.34	5.92	0.85	0.86	0.60	0.77
0.30	0.36	-0.10	-1.42	-4.56	-1.03	0.60	1.14	1.06
0.18	0.58	0.35	-1.96	-5.27	-1.73	0.77	1.06	0.65

(a)

0.48	1.57	1.66	-3.94	-9.14	-3.56	2.34	2.40	1.19
0.52	0.72	0.26	3.35	-3.11	1.46	2.77	2.86	1.44
-0.18	-1.49	-0.56	10.74	17.32	10.49	2.57	1.83	1.23
-3.05	3.11	10.72	-24.13	-5.05	-6.62	3.86	1.10	0.42
-5.85	-2.97	14.47	-11.60	100.00	22.01	2.05	-5.23	-2.11
-2.76	-1.76	-1.72	-9.66	-13.02	9.92	-5.34	-1.18	-0.18
-0.80	-4.97	-4.00	8.07	34.45	7.24	-1.77	-3.35	0.31
0.24	-3.10	-4.68	5.30	11.77	2.95	-2.97	-2.48	0.20
0.67	-0.78	-1.99	1.00	1.79	0.56	-2.78	-1.69	-0.09

(b)

4. Restoration Example: Deblurring

The wavelet quantization noise model introduced in previous sections has been incorporated in several restoration algorithms, including image deblurring and multi-frame motion imagery restoration. Although Section 2 introduced several different types of probability distribution functions that could be used, experiments have shown that the Gaussian noise model that makes use of the full $\mathbf{K}_{\mathbf{e_z}}$ covariance matrix gives the best restoration results. In order of increasing performance, the following distribution types have been investigated with regard to their restoration effectiveness: IID Laplacian (distribution is a product of univariate Laplacian distributions); independent Laplacian distribution with variances taken from diag($\mathbf{K}_{\mathbf{e_z}}$); IID Gaussian; independent Gaussian with variances taken from diag($\mathbf{K}_{\mathbf{e_z}}$); the multivariate Laplacian described by (9); and the Gaussian noise model with covariance matrix $\mathbf{K}_{\mathbf{e_z}}$. The latter two of these distributions give nearly identical restoration results, which can be explained by the similarity in the gradients of the log-likelihoods of the two distributions. Section 5, which provides more-extensive simulations with different noise models within a temporal filtering framework, corroborates this informal ranking of the noise models' accuracy.

The remainder of this section is devoted to the particular restoration case of deblurring wavelet-compressed images. Although different types of blurring are possible, this section will focus on linear motion blurs. A brief problem formulation will be presented, followed by experimental results that demonstrate the advantage of using the Gaussian quantization noise model compared to using an IID Gaussian noise model. Since the results corresponding to the Laplace noise models are inferior to those of the Gaussian noise models, the formulations for the Laplacian case are omitted; nevertheless, such formulations are straightforward and follow very closely the developments that follow.

4.1. Problem Formulation

If the original image is denoted as $\mathbf{z}$, the observation model for the blurred image $\mathbf{z}_b$ can be written as

$$\mathbf{z}_b = \mathbf{B}\mathbf{z} + \mathbf{n}_a, \tag{11}$$

where $\mathbf{B}$ is the matrix representing the motion blur, and $\mathbf{n}_a$ represents acquisition noise of the imaging system. While the observation vector $\mathbf{z}_b$ is of size $N \times 1$, the original image vector $\mathbf{z}$ is of size $M \times 1$, $M > N$, due to the influence of image content beyond the border of the observed image; the blurring matrix $\mathbf{B}$ is $N \times M$. We do not include any nonlinearities in the observation model (due to, for example, the response characteristics of film or gamma processing), although alternative formulations that include them are possible. Note that there are a number of methods that are commonly used for handling the boundary conditions in the deblurring problem, including zero-padding (or padding with some other constant) outside the boundary, periodic extension beyond the boundary, and symmetric extension beyond the boundary; however, while some of these alternatives may offer computational incentive for their use, none model the actual blurring process as accurately as the method described above which has the number of pixels to be estimated larger than the number of observations, i.e., $M > N$.

As described in previous sections, the image $\mathbf{z}_b$ can be transformed by a DWT, quantized, and then inverse transformed back to the spatial domain to form the observed image $\mathbf{z}_q$,

$$\begin{aligned} \mathbf{z}_q &= \mathbf{z}_b + \mathbf{e}_\mathbf{z} \\ &= \mathbf{B}\mathbf{z} + \mathbf{n}_a + \mathbf{e}_\mathbf{z}, \end{aligned} \tag{12}$$

where the quantization noise $\mathbf{e}_\mathbf{z}$ is equivalent to the noise term discussed in previous sections. For many situations, the quantization noise $\mathbf{e}_\mathbf{z}$ will dominate the acquisition noise $\mathbf{n}_a$ (compression has a tendency to smooth out and remove much of this noise), and we will approximate (12) as

$$\mathbf{z}_q = \mathbf{B}\mathbf{z} + \mathbf{e}_\mathbf{z}. \tag{13}$$

Although not implemented for the deblurring case, the methods of Section 5.4 could be used to model explicitly the combined effect of $\mathbf{n}_a$ and $\mathbf{e}_\mathbf{z}$.

It is well known that care must be taken when estimating $\mathbf{z}$ from its blurred and noisy observation $\mathbf{z}_q$, for otherwise excessive noise amplification can result [19]. To avoid such problems, the restoration here makes use of a prior image model that discourages noise amplification, and solves the problem in a Bayesian framework. In particular, the image

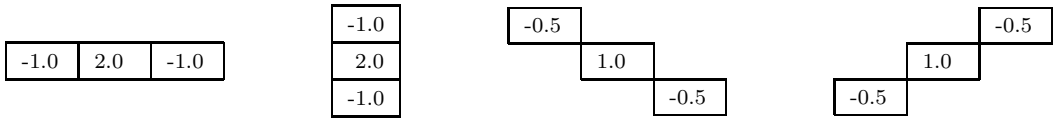


Fig. 8. Masks used for construction of $\mathbf{d}_c^t\mathbf{z}$. For example, for the first of the four cliques above, $\mathbf{d}_c^t\mathbf{z} = -Z[m,n-1] + 2Z[m,n] - Z[m,n+1]$, where $Z[m,n]$ is the $(m,n)^{th}$ element of the image represented by $\mathbf{z}$.

estimate is taken as the maximum a posteriori solution that maximizes the posterior distribution of $\mathbf{z}$ given $\mathbf{z}_q$, which is equivalent to maximizing $p(\mathbf{z})p(\mathbf{z}_q|\mathbf{z})$. The likelihood term $p(\mathbf{z}_q|\mathbf{z})$ is found by combining (6) and (13),

$$p(\mathbf{z}_q|\mathbf{z}) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{B}\mathbf{z}-\mathbf{z}_q)^t\mathbf{K}_{\mathbf{e_z}}^{-1}(\mathbf{B}\mathbf{z}-\mathbf{z}_q)\right\}}{(2\pi)^{N/2}\left|\mathbf{K}_{\mathbf{e_z}}\right|^{1/2}}. \tag{14}$$

A Markov Random Field (MRF) [20] is used to describe the prior image probability $p(\mathbf{z})$,

$$p(\mathbf{z}) = \frac{1}{V}\exp\left\{-\lambda\sum_{c\in\mathcal{C}}\rho_T\left(\mathbf{d}_c^t\mathbf{z}\right)\right\}, \tag{15}$$

where V is a normalizing constant, λ is related to the temperature parameter associated with the Gibbs prior distribution [20], c are indexes of local groups of pixels called cliques, and $\mathcal{C}$ is the overall set of such cliques. The $\mathbf{d}_c^t\mathbf{z}$ terms form spatial activity measures which effectively discourage the type of noise amplification possible when deblurring images. Here, the $\mathbf{d}_c^t\mathbf{z}$ terms at each pixel location are chosen to approximate second-order directional derivatives, with coefficients as shown in Figure 8. The function $\rho_T(\cdot)$ controls how heavily the spatial activity measures are penalized, and is chosen here as

$$\rho_T(\varepsilon) = \begin{cases} \varepsilon^2 & |\varepsilon| \le T \\ T^2 + 2T(|\varepsilon| - T) & |\varepsilon| > T \end{cases}, \tag{16}$$

shown graphically in Figure 9 for $T = 4$. Such a choice of penalizing function results in a Huber-Markov Random Field (HMRF), which has been used with much success in other image and video restoration problems [21][22]. The $\rho_T(\cdot)$ function quadratically penalizes values less than or equal to T, while linearly penalizing values greater than T, which prevents sharp image discontinuities like edges from being excessively blurred as can happen when using a purely-quadratic penalty function. For further detail on the use of the HMRF in image restoration, the reader is referred to [21][22][23].

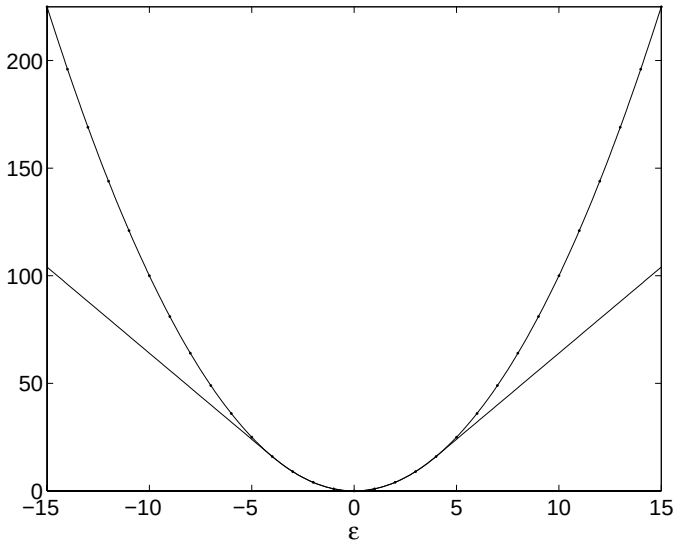


Fig. 9. The Huber penalty function for $T = 4$ (solid) compared with a quadratic penalty function (dotted).

Since maximization of the posterior distribution is equivalent to minimization of the negative of its natural logarithm, the image estimate can be written as the solution to

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \frac{1}{2}(\mathbf{B}\mathbf{z} - \mathbf{z}_q)^t \mathbf{K}_{\mathbf{e_z}}^{-1}(\mathbf{B}\mathbf{z} - \mathbf{z}_q) + \lambda \sum_{c \in \mathcal{C}} \rho_T \left(\mathbf{d}_c^t \mathbf{z}\right). \tag{17}$$

The solution to this convex minimization problem is determined iteratively based on a gradient descent algorithm, which starts with an initial estimate $\mathbf{z}_{(0)}$ and updates the estimate as

$$\mathbf{z}_{(i+1)} = \mathbf{z}_{(i)} - \alpha_{(i)} \mathbf{g}_{(i)}, \tag{18}$$

where $\mathbf{g}_{(i)}$ is the gradient of the objective function in (17) evaluated at $\mathbf{z}_{(i)}$, and $\alpha_{(i)}$ is a step size which controls how far the estimate for iteration $(i+1)$ moves from $\mathbf{z}_{(i)}$ along $\mathbf{g}_{(i)}$. The initial estimate $\mathbf{z}_{(0)}$ is taken as $\mathbf{z}_q$ for the N observed pixels; the $M - N$ off-screen pixels of $\mathbf{z}_{(0)}$ are computed by blurring $\mathbf{z}_q$ in the direction opposite of the blur $\mathbf{B}$.

The gradient is computed from (17) as

$$\begin{aligned} \mathbf{g}_{(i)} &= \mathbf{B}^t \mathbf{K}_{\mathbf{e_z}}^{-1}(\mathbf{B}\mathbf{z}_{(i)} - \mathbf{z}_q) + \lambda \sum_{c \in \mathcal{C}} \mathbf{d}_c \rho_T'(\mathbf{d}_c^t \mathbf{z}_{(i)}) \\ &= \mathbf{B}^t \mathbf{H}^t \mathbf{K}_{\mathbf{e_y}}^{-1} \mathbf{H}(\mathbf{B}\mathbf{z}_{(i)} - \mathbf{z}_q) + \lambda \sum_{c \in \mathcal{C}} \mathbf{d}_c \rho_T'(\mathbf{d}_c^t \mathbf{z}_{(i)}). \end{aligned} \tag{19}$$

For real unitary discrete transforms (for example, the DCT), $\mathbf{H}^t = \mathbf{H}^{-1}$, and the term $\mathbf{H}^t$ above is easily implemented as the inverse transform. Since all DWT matrices of practical interest for image compression are not unitary, one must implement a new transform operation that is equivalent to $\mathbf{H}^t$. Appendix A discusses the implementation of the transpose of the DWT matrix.

The step size to be used in conjunction with the gradient is given as

$$\alpha_{(i)} = \frac{\mathbf{g}_{(i)}^t \mathbf{g}_{(i)}}{\mathbf{g}_{(i)}^t \left[\mathbf{B}^t \mathbf{H}^t \mathbf{K}_{\mathbf{e_y}}^{-1} \mathbf{H} \mathbf{B} + \lambda \sum_{c \in \mathcal{C}} \mathbf{d}_c \rho_T''(\mathbf{d}_c^t \mathbf{z}_{(i)}) \mathbf{d}_c^t \right] \mathbf{g}_{(i)}}, \tag{20}$$

whose derivation is given in Appendix B. Iterations of (18) are continued until the convergence criterion

$$\frac{\left| \mathcal{O}\left(\mathbf{z}_{(i+1)}\right) - \mathcal{O}\left(\mathbf{z}_{(i)}\right) \right|}{\mathcal{O}\left(\mathbf{z}_{(i)}\right)} < \epsilon \tag{21}$$

is met, where $\mathcal{O}(\cdot)$ represents the objective function in (17) and ϵ is a small threshold that sets the rate of decrease in the objective function that is required for convergence. If the number of iterations exceeds some pre-determined threshold (taken as 1500 here) without convergence according to (21), the algorithm is also terminated.

4.2. Experimental Results

To test the deblurring algorithm, the 512×512 grayscale image *boat* image is first blurred with a filter $\mathbf{B}$ that simulates the effect of the spatial and temporal integration of a CCD image array when the relative motion between the scene and camera over the integration time of the sensor has a distance of 5.7 pixels at angle $-\pi/9$, where the velocity is constant. The central 384×384 pixels of the result are then taken as the blurred image. (Using the central region like this better simulates real data, because these 384×384 pixels contain information from other pixels outside of the sub-image, as would be expected from a true camera.) The coefficients of a four-level 384×384 Daubechies 9/7 wavelet decomposition are quantized as described in Section 3, and the final observed image is taken as the inverse wavelet transform of the dequantized coefficients; reconstruction levels are taken as the midpoints of the quantization cells. The PSNR of the de-compressed blurred image relative to the uncompressed blurred image is 42.8 dB.

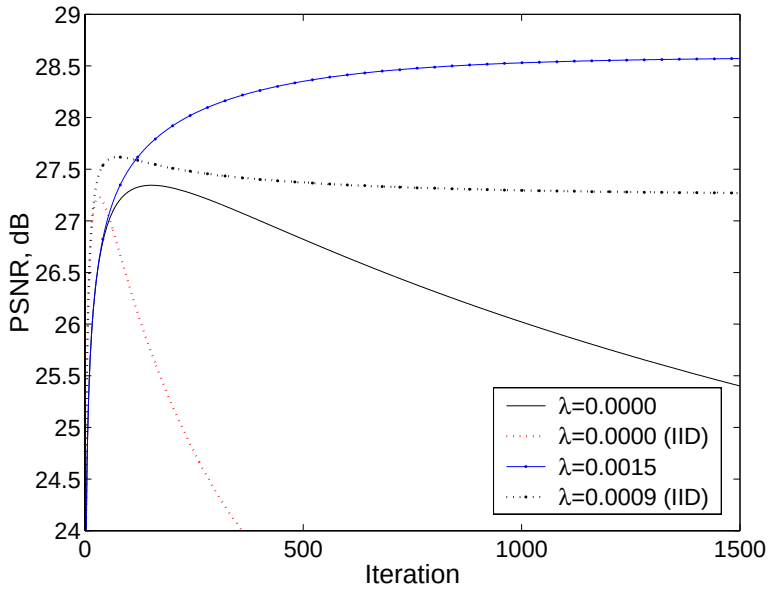


Fig. 10. Plots of PSNR vs. iterations for deblurring. Examples are given for maximum likelihood restorations ($\lambda = 0.0$), for both IID and the proposed noise models; and for MAP restorations ($\lambda > 0.0$), for both IID and the proposed noise model.

Two methods of deblurring are presented. The first method uses the wavelet quantization noise model as described in Section 4.1, while the second method replaces the noise covariance term $\mathbf{K}_{\mathbf{e_z}}^{-1}$ in (17) with a diagonal matrix consisting of elements $12/\Delta_d^2$, where Δ_d was introduced in Section 3. The latter of these two methods implies independent and identically distributed Gaussian noise, which for the case at hand leads to a deblurring estimate that makes use of (18) and modified versions of (19) and (20), omitted here for brevity. For the IID model, variances of $\Delta_d^2/12$ provide rough estimates of the spatial-domain quantization noise. For all cases reported here, the Huber parameter was chosen as $T = 4.0$.

The quality metric used here is PSNR, defined as $10\log_{10}(255^2/MSE)$, where MSE is the mean-squared error between a reconstruction and the reference image. Figure 10 plots PSNR as a function of iteration for several values of λ, including the case of $\lambda = 0.0$ which removes the regularizing smoothness term from the formulation to form a maximum likelihood estimate. From the figure, it is readily evident that there are two PSNR values which could be used for comparing the two methods—the converged value, PSNR_c, which is

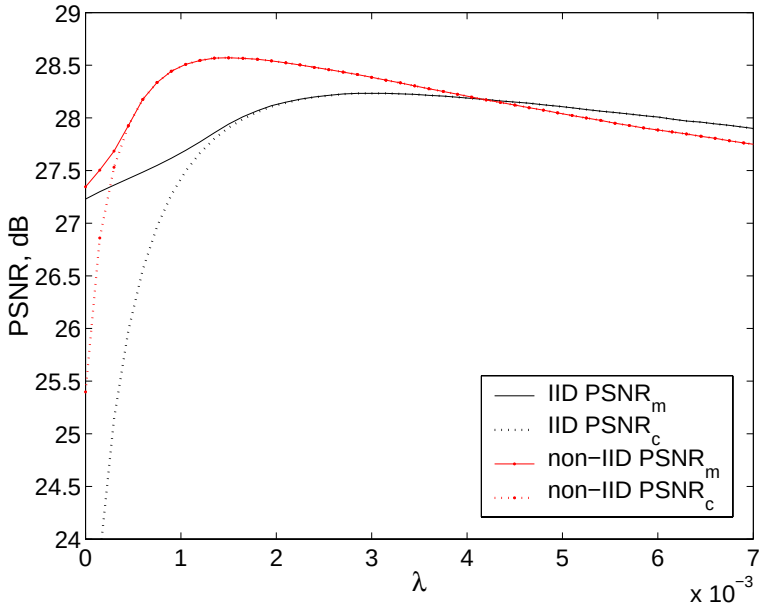


Fig. 11. Plots of λ vs. PSNR for deblurring, showing both the PSNR at the peak of the plots from Figure 10 and at the converged solutions.

the PSNR achieved after 1500 iterations or convergence; and the maximum value, PSNR_m, which is the maximum PSNR achieved and may or may not be equal to PSNR_c. In an actual deblurring scenario, it is impossible to know exactly when the maximum PSNR is achieved, but careful monitoring of deblurring progress would allow a user to terminate an algorithm when the reconstruction appears best. Thus PSNR_m is a valid measure, although a hands-off reconstruction (where the estimate is taken at algorithm convergence, hence using PSNR_c) would be preferred.

Figure 11 plots the two quality measures PSNR_c and PSNR_m as a function of λ. The best reconstruction of the model making use of the quantization noise model is about 0.26 dB better than the best reconstruction using the IID noise model. Results for other test images under similar conditions are shown in Table 3. While these improvements are not large, they do demonstrate that using the quantization noise model consistently provides reconstructions that are mathematically closer to the original than when using the IID noise model. For all simulations performed, including those not reported in Table 3, PSNR values of the

TABLE 3

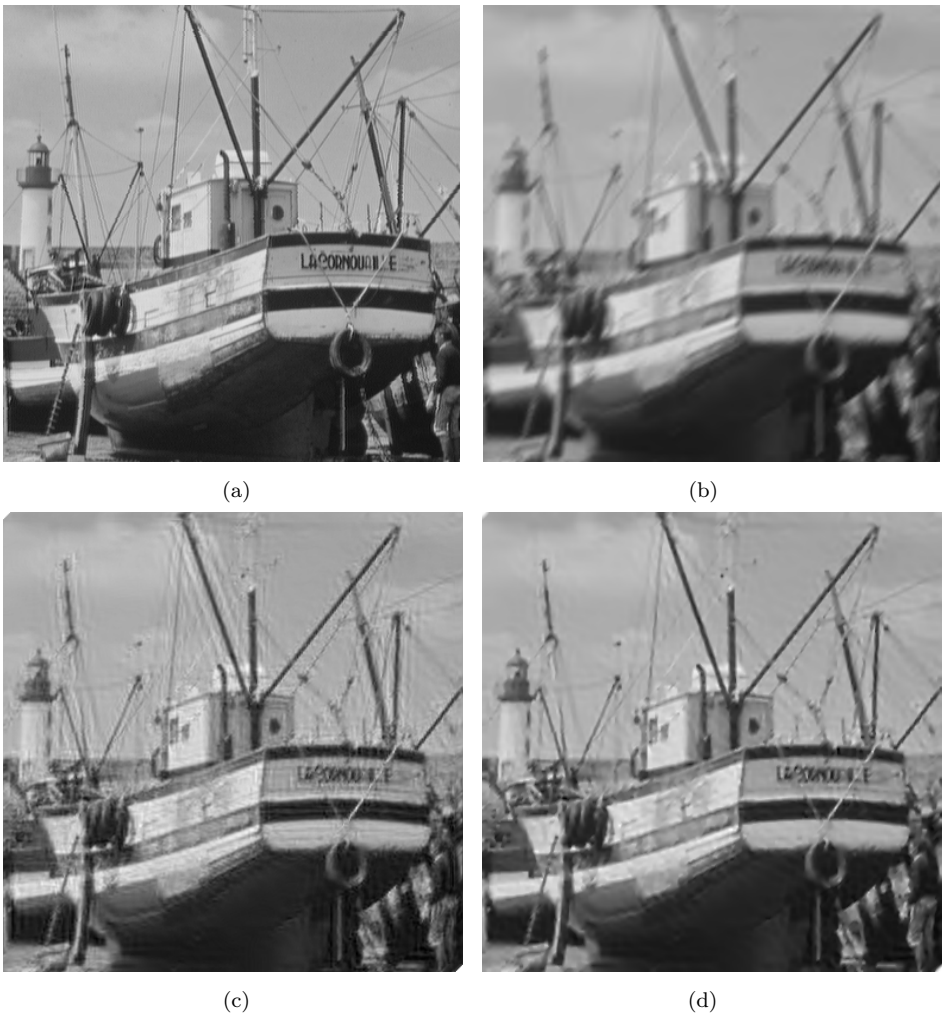
PSNR improvements for restoration using quantization noise model versus restoration assuming IID noise model.

sequence	Compressed PSNR	Motion Length	Motion Angle	ΔPSNR_c	ΔPSNR_m
barb	46.4	7.5	$\pi/7$	0.48	0.36
boat	43.2	9.0	$\pi/4$	0.34	0.34
boat	42.8	5.7	$-\pi/9$	0.26	0.26
goldhill	42.6	10.0	$-\pi/6$	0.22	0.21
goldhill	41.2	5.0	$-\pi/3$	0.19	0.19
lena	41.8	11.1	$\pi/5$	0.37	0.37
mandrill	38.4	14.2	$\pi/5$	0.16	0.14
mandrill	39.0	4.1	$-4\pi/5$	0.22	0.23
peppers	46.2	12.1	$4\pi/5$	0.36	0.36

reconstructions were higher when using the quantization noise model than when using the IID noise model. Since the only difference between the two restoration methods is in their noise characterization, it stands to reason that the quantization noise model of Section 2 provides a more accurate description of the compression error than does an IID noise model.

Visual results of the reconstructions that use the quantization noise model are given in Figure 12. Part (c) of the figure shows the reconstruction for $\lambda = 0.0$, which corresponds to the maximum likelihood solution; part (d) shows the MAP reconstruction. For both reconstructions shown, the estimate is taken at the iteration that maximizes PSNR. The ML estimate was thus taken at the peak of the ML plot shown in Figure 10, while the MAP estimate had its highest PSNR at convergence. Superiority of the MAP estimate to the ML estimate is apparent in the figure, which demonstrates the noise suppression introduced by the prior model. Distinguishing between the reconstructions of the IID noise model and the quantization noise model is difficult on the printed page, and thus visual results for the IID noise assumption are not presented here.

As a side note, the reconstructions in the figure have additional spatial extent than the 384×384 original (due to the spatial blurring at the image boundaries with pixels outside of the field of view), which may or may not be important to a user of a deblurring application; in either case, using additional pixels outside of the boundary helps to prevent excessive ringing near the borders, as might be expected when using the less-accurate boundary conditions



(a) (b)

(c) (d)

Fig. 12. Visual deblurring results when using the quantization noise model. The original image is shown in (a), with the blurred and compressed image shown in (b). The PSNR of the blurred and compressed image in (b) relative to the blurred but uncompressed image is 43.2 dB, while the PSNR of the image in (b) relative to the image in (a) is 20.0 dB. Results are all taken at the point of best reconstructed PSNR. (c) Maximum likelihood estimate, $\lambda = 0.0$, $\text{PSNR}_m = 27.34$ dB; (d) MAP estimate, $\lambda = 1.5 \times 10^{-3}$, $\text{PSNR}_m = 28.57$ db.

mentioned previously. Although not reported in Table 3, it is interesting to note that using the Huber penalty function $\rho_T(\cdot)$ gives significant improvement relative to using a quadratic penalty function[2] corresponding to a Gauss-Markov Random Field (GMRF). Discounting the noise model, the HMRF gave up to 1.2 dB improvement relative to the GMRF for the

[2]A quadratic penalty function corresponds to allowing the Huber parameter T to become arbitrarily large.

cases presented above.

5. Restoration Example: Temporal Filtering

This section demonstrates various methods of restoring motion imagery compressed by scalar quantization of both 2D and 3D wavelet coefficients. In both cases, the data of interest are motion imagery, i.e., they are sequences of images indexed by time. Filtering along the time domain allows the incorporation of information from surrounding frames for the restoration of the current frame. Such ideas have been used in the super-resolution literature (for example, in [22]), where the goal is to improve the spatial resolution by using information from surrounding frames; here, however, the goal is to improve the quality of a frame without changing resolution. Several examples of temporal filtering in this manner can be found in the literature [24][25][26][27]. The section begins by formulating the general framework for temporal filtering, and continues by examining temporal filtering when using several different variations of the quantization noise model discussed in this report. The first two subsections provide results for synthetically-generated video whose individual frames have been compressed by using the 2D DWT. Subsection 5.3 extends these results to the case of compression with a 3D DWT. Subsection 5.4 considers the case of actual video, showing that benefits of using the proposed quantization noise model apply for real video as well the synthetic sequences used in the first three subsections.

Suppose the original image at time k is $\mathbf{z}^k$. Since a temporal filter uses pixel information from frames at different time instants, one must first describe the relationships between the images at different times. One common method of relating images at two time instants k and l is to model them such that

$$\mathbf{z}^l = \mathbf{A}_{k,l}\mathbf{z}^k + \boldsymbol{\mu}^{k,l} + \mathbf{e}^{k,l}, \tag{22}$$

where the matrix $\mathbf{A}_{k,l}$ forms a prediction of $\mathbf{z}^l$ given $\mathbf{z}^k$, $\mathbf{e}^{k,l}$ is the error in such a prediction, and $\boldsymbol{\mu}^{k,l}$ is a "mean" term that accounts for pixels in $\mathbf{z}^l$ that are unobservable from $\mathbf{z}^k$. Note that when $l = k$, $\mathbf{A}_{k,k} = \mathbf{I}$, $\mathbf{e}^{k,k} = \mathbf{0}$, and $\boldsymbol{\mu}^{k,k} = \mathbf{0}$. When the original images are compressed, the model that relates the observation at time l to the original image at time k must be modified according to the quantization error,

$$\mathbf{z}_q^l = \mathbf{A}_{k,l}\mathbf{z}^k + \boldsymbol{\mu}^{k,l} + \mathbf{e}^{k,l} + \mathbf{e}_{\mathbf{z}}^l, \tag{23}$$

where the additional noise term is the same quantization error term introduced in Section 2. For notational convenience, the two noise terms can be combined into a single noise term,

$$\mathbf{z}_q^l = \mathbf{A}_{k,l}\mathbf{z}^k + \boldsymbol{\mu}^{k,l} + \mathbf{n}^{k,l}. \tag{24}$$

The error is assumed to be normally distributed, with mean $\mathbf{0}$ and covariance $\mathbf{K}_{k,l}$, which yields the probability distribution function of $\mathbf{z}_q^l$ given $\mathbf{z}^k$,

$$p(\mathbf{z}_q^l|\mathbf{z}^k) = \frac{1}{(2\pi)^{N/2}|\mathbf{K}_{k,l}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{A}_{k,l}\mathbf{z}^k + \boldsymbol{\mu}^{k,l} - \mathbf{z}_q^l)^t \mathbf{K}_{k,l}^{-1}(\mathbf{A}_{k,l}\mathbf{z}^k + \boldsymbol{\mu}^{k,l} - \mathbf{z}_q^l)\right\}. \tag{25}$$

With the simplifying approximation that $\mathbf{z}_q^l|\mathbf{z}^k$ and $\mathbf{z}_q^m|\mathbf{z}^k$ are independent for $m \neq l$, the joint pdf of observations of compressed images at times $k-n, \ldots, k+n$ is

$$p(\mathbf{z}_q^l|\mathbf{z}^k, l = k-n, \ldots, k+n) = \prod_{l=k-n}^{k+n} p(\mathbf{z}_q^l|\mathbf{z}^k). \tag{26}$$

The temporal filters described in this section all find maximum likelihood estimates of a single frame $\mathbf{z}^k$ given degraded observations of the sequence at time instants within $\pm n$ of k, i.e., $k-n, \ldots, k+n$. The maximum likelihood estimator chooses an estimate for $\mathbf{z}^k$ that maximizes the likelihood term in (26). Since maximizing a function is equivalent to minimizing the negative of its natural logarithm, the ML estimate can be written as

$$\hat{\mathbf{z}}^k = \arg\min_{\mathbf{z}^k} \sum_{l=k-n}^{k+n} (\mathbf{A}_{k,l}\mathbf{z}^k + \boldsymbol{\mu}^{k,l} - \mathbf{z}_q^l)^t \mathbf{K}_{k,l}^{-1}(\mathbf{A}_{k,l}\mathbf{z}^k + \boldsymbol{\mu}^{k,l} - \mathbf{z}_q^l). \tag{27}$$

The various temporal filtering schemes presented in this section differ in their dimensionality (two- or three-dimensional DWT) as well as their choice of covariance matrix $\mathbf{K}_{k,l}$, but the general problem setup is that of (27). Equation (27) is solved using an iterative conjugate gradient optimization algorithm, details of which are given in Appendix C. Note, however, that some of the simplified noise terms lead to simplifications that allow almost trivial implementations that are not iterative; such simplifications will be discussed as they are introduced in later subsections.

Elements of the mean term $\boldsymbol{\mu}^{k,l}$ are chosen as

$$\boldsymbol{\mu}^{k,l} = \begin{cases} 0 & \text{for observable pixels} \\ \mathbf{z}_q^l & \text{for unobservable pixels} \end{cases}. \tag{28}$$

For pixels at time l that do not correspond to any pixels at time k, the corresponding rows of $\mathbf{A}_{k,l}$ consist entirely of zeros. To classify a pixel at time l as having or not having

corresponding pixels at time k, a simple threshold can be applied to the difference $|\mathbf{z}_q^l - \mathbf{A}_{k,l}\mathbf{z}_q^k|$ after motion estimation; knowledge of pixels appearing or disappearing at image edges due to camera motion can also be used. More sophisticated algorithms could be used to estimate both $\mathbf{A}_{k,l}$ and $\boldsymbol{\mu}^{k,l}$ simultaneously, although they are not explored here.

The following subsections discuss various methods and experiments of temporally filtering wavelet-compressed motion imagery.

5.1. Experiment 1

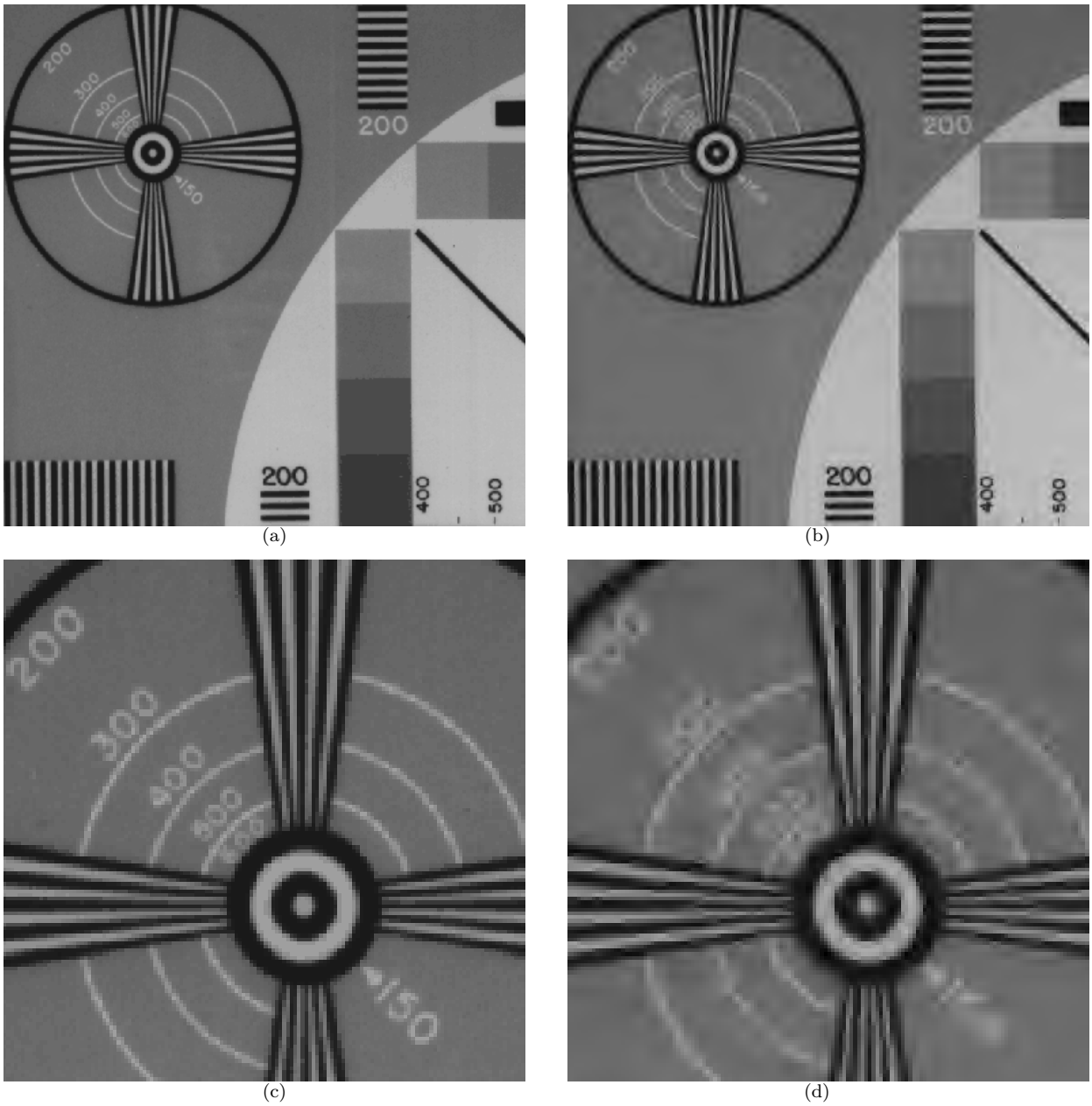
Consider the case where each frame of an image sequence is compressed by use of a 2D DWT, independently of the other frames. This subsection considers the simplified case where both the horizontal and vertical motion between frames is an integer number of pixels. The simplest of temporal filters assumes Gaussian noise with covariance matrix $\mathbf{K}_{k,l}$ that is diagonal with elements σ^2 for each $l = k-n, \ldots, k+n$. Such a filter equally weights all observations of each frame at each time instant, and can be loosely interpreted as a box filter in the temporal dimension along the motion trajectories described by the $\mathbf{A}_{k,l}$ matrices. Another simple filter assumes IID Laplace noise, and results in a median filter along the motion trajectories. More accurate versions of these two filters can be achieved by using the same noise type (i.e., independent Gaussian or Laplacian), but assigning $\mathbf{K}_{k,l} = \mathrm{diag}(\mathbf{K}_{\mathbf{e}_\mathbf{z}^l})$. For the Gaussian case, this leads to a weighted-average filter along motion trajectories such that pixels that are more accurate according to the quantization noise model are weighted more heavily than pixels that are less accurate. Similarly, the Laplace case leads to a weighted-median filter along motion trajectories. The fifth and final case considered here is to use Gaussian noise with $\mathbf{K}_{k,l} = \mathbf{K}_{\mathbf{e}_\mathbf{z}^l}$, which cannot be implemented by simple averaging or median filtering, but is instead implemented with the conjugate gradient method discussed in Appendix C. All five of these noise models for $\mathbf{K}_{k,l}$ consider only the noise introduced by compression.

The initial experiment given in this subsection is designed to isolate the effect of the quantization noise model. The input sequence consists of five frames, where each frame is taken from a single source image and globally translated with integer shift; thus the only difference between these five images is that they are global translates of each other, where

the global motion parameters are known. In a very simple sense, these five images form an artificially-constructed "image sequence." Obviously, sequences with such a property will almost never occur in actual video; however, the point of this subsection is to isolate the effect of the quantization noise model, and the method discussed here eliminates the influence of other motion-related error such as $\mathbf{e}^{k,l}$ or uncertainty in the matrices $\mathbf{A}_{k,l}$. Subsection 5.4 will relax the constraints imposed here so as to filter actual video.

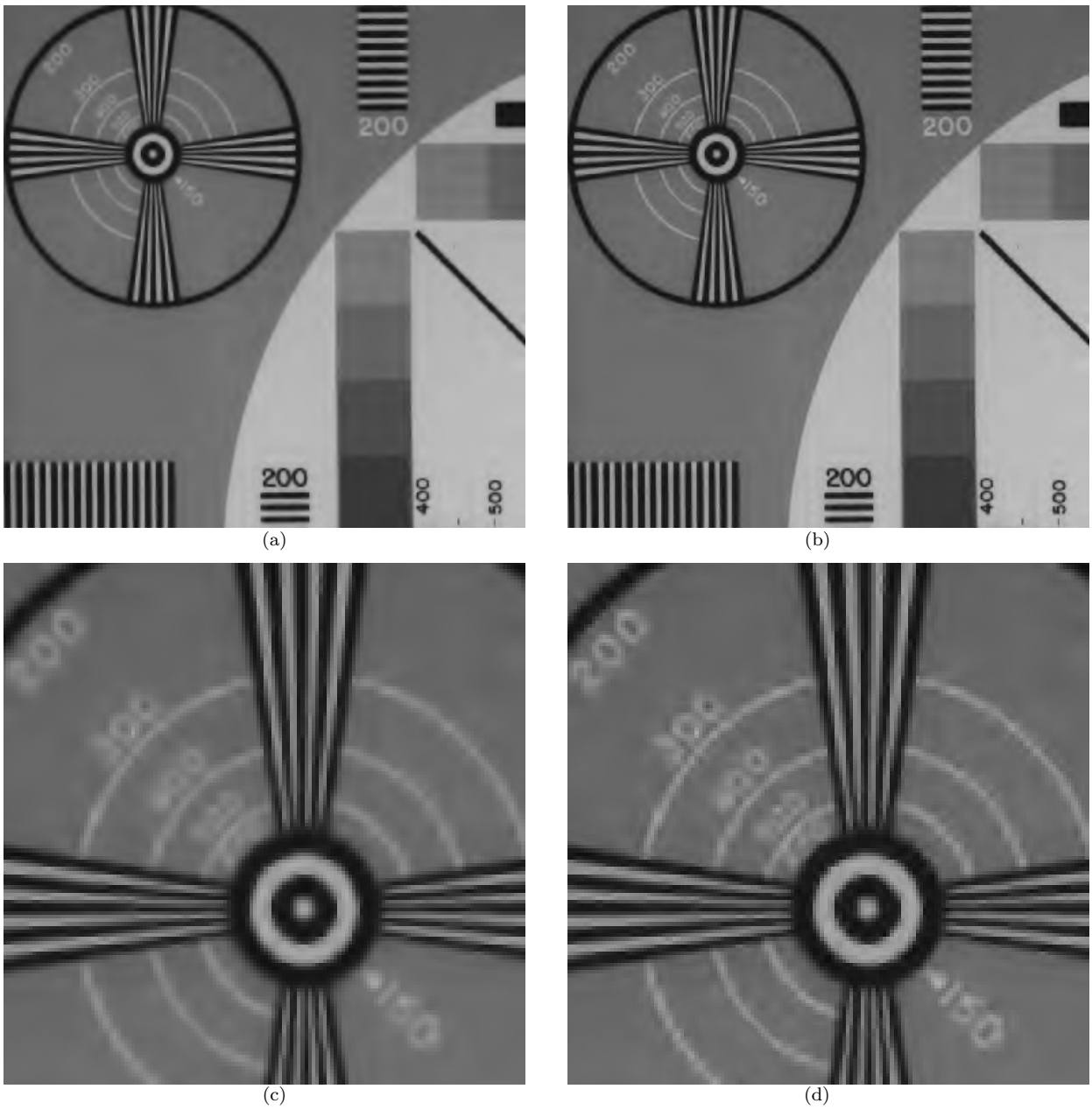
The integer shifts that relate the five frames are $(0,0)$, $(1,0)$, $(0,1)$, $(-1,0)$, and $(0,-1)$. The compressed images are formed by quantization of four-level DWT decompositions of the images, where the quantization is performed as described by (10). The frames used for compression are 384×384 portions of the original larger image, which allows the shifted frames to be extracted without missing pixels at the image borders. Figure 13 shows the original *test-pattern* image along with the image to be restored, the $(0,0)$-shifted frame. Although only the compressed image corresponding to $(0,0)$ translation is shown in the figure, the other four compressed images appear approximately the same; all five compressed images for *test-pattern* have PSNR in the range 32.84 ± 0.06 dB.

Figure 14 compares the restoration results of *test-pattern* for the IID Gaussian noise model and for the proposed noise model. As can be seen by comparing Figures 13 and 14, even the IID Gaussian noise model yields considerable improvement over the originally-compressed image; this, however, is to be expected due to the contrived nature of the experiment. The important lesson to be learned from Figure 14 is not the improvement relative to the image in Figure 13(b), but rather the improvement of the images in Figure 14 (b) and (d) relative to those in parts (a) and (c)—using the quantization noise model proposed in this technical report provides over 1 dB of improvement in PSNR relative to using an IID Gaussian noise model, and, most importantly, there is a significant visible improvement as well. Table 4 summarizes results for other test images using all five quantization noise models introduced at the beginning of this subsection; the 384×384 portions of each of these test images are shown in Figure 15. As can be seen from Table 4, using the full quantization error covariance matrix with a Gaussian pdf produces PSNR results that are significantly better than those produced when using IID distributions or when using reduced versions of the full covariance matrix, and is true for either Laplace or Gaussian distributions.



(a) (b) (c) (d)

Fig. 13. Original and compressed images for the first temporal filtering experiment: (a) original; (b) compressed image 0, PSNR=32.80 dB; (c) close-up of original; (d) close-up of compressed image.



(a) (b)

(c) (d)

Fig. 14. Comparison of restorations of the compressed image in Figure 13 using IID non-IID Gaussian noise models. (a) Restoration using IID Gaussian noise model, PSNR=35.10 dB; (b) restoration using Gaussian noise model with covariance matrices $\mathbf{K}_{\mathbf{e}_{\mathbf{z}}^l}$, PSNR=36.14 dB; (c) close-up of IID Gaussian case; (d) close-up of non-IID Gaussian noise case.

TABLE 4

PSNR values for restoration using varous quantization noise models. The range of PSNR's under "Compressed images" refers to the PSNR's for the five compressed input images. All PSNR values are in decibels.

Sequence name	Compressed images	Laplacian, IID	Laplacian, diag($\mathbf{K}_{\mathbf{e}_z^l}$)	Gaussian, IID	Gaussian, diag($\mathbf{K}_{\mathbf{e}_z^l}$)	Gaussian, $\mathbf{K}_{\mathbf{e}_z^l}$
barb	32.86 ± 0.02	35.18	35.50	36.00	36.31	37.15
boat	30.49 ± 0.03	32.06	32,24	32.65	32.84	33.71
mandrill	32.29 ± 0.01	34.79	35.04	35.52	35.87	36.83
peppers	32.50 ± 0.03	33.48	33.55	33.81	33.90	34.37
test-pattern	31.10 ± 0.06	32.42	32.61	33.26	33.42	34.24
test-pattern	32.84 ± 0.06	34.30	34.51	35.10	35.29	36.14
test-pattern	34.73 ± 0.07	36.39	36.62	37.10	37.31	38.09

(a) (b) (c) (d)

Fig. 15. Original input images. (a) mandrill; (b) boat; (c) peppers; (d) barb.

The five test images used here contain various characteristics, from the text contained in the resolution chart of *test-pattern*, to the texture of *mandrill*'s fur or *barb*'s clothing, to the smooth regions contained in areas of each of the images. Although the results presented in Table 4 do not contain an exhaustive comparison between images at different compression qualities, the PSNR improvements evident in the table are representative of PSNR improvements that have been observed over a wide range of compression severities.

5.2. Experiment 2

The previous subsection demonstrated the potential improvement of using the proposed quantization noise model relative to using simple IID Gaussian or Laplacian noise models. Both the proposed noise model and the IID noise models provided significant improvements

relative to the compressed images, with the proposed noise model giving approximately an additional dB of improvement over the IID Gaussian model. The experiment presented in this subsection will demonstrate that the improvements that one gets, for each of the noise models, is quite dependent on the actual motion between the frames.

Consider an experiment much like the one of the previous subsection, in which original frames of a synthetic video sequence are formed by taking 384×384 subimages from an original 512×512 image; each of the frames are offset by an integer global translation. In this subsection, the frame to be restored is considered as having global translation of $(0, 0)$, and eight other frames are formed by taking shifts of $\pm b$ pixels, i.e., $(b, 0)$, (b, b), $(0, b)$, $(-b, b)$, $(-b, 0)$, $(-b, -b)$, $(0, -b)$, and $(b, -b)$. Table 5 compares PSNR improvements as a function of b for Gaussian noise models making use of $\mathbf{K}_{\mathbf{e}_{\mathbf{z}}^l}$ and the IID noise assumption; the *mandrill* image is used for testing. For this simple restoration example, the PSNR improvements obviously depend on the motion that occurs between frames—odd pixel displacements consistently yield better results than even displacements. These results in the table are quite interesting; for each value of b, the input frames all have (approximately) the same PSNR. Thus, although in each case of b the input images are shifted versions of each other with equivalent PSNR's, there are drastic differences in PSNR results when the samples are averaged. Such behavior is easily explained: Due to the subsampling employed by a non-expansive discrete wavelet transform, shifts of $b = 2$ among frames result in the highest-frequency subbands' being exact shifts of each other. Since the quantization parameters are unchanged between images, the quantized DWT coefficients in these highest-frequency subbands are identical among the different images, and hence no new information is introduced in the highest subbands of the shifted image observations. Any gains for $b = 2$ are due entirely to lower-frequency subbands. Similarly, for the case of $b = 4$ not only are the highest-frequency subbands exact shifts of each other, but so are the second-highest-frequency subbands; hence the PSNR improvements for $b = 4$ are worse than for $b = 2$. The trend continues for $b = 8$, and ultimately one can conclude that for a lev-level DWT decomposition, shifts of $\pm v 2^{lev}$, v an integer, there will be zero gain by filtering multiple observations; such is the case regardless of the noise model one uses. An analogous phenomenon occurs when a compression technique makes use of the block DCT: If the block

TABLE 5

Dependence of filtering results on integer shifts b for the *mandrill* image. Here, frame 0 is being restored based on the nine total observations that are within $\pm b$ of a shift of $(0, 0)$. Quantized input frames all have PSNR of 34.1 ± 0.05 dB.

b	ΔPSNR, Gaussian IID	ΔPSNR, Gaussian $\mathbf{K}_{\mathbf{e}_z^l}$
1	3.93	5.69
2	0.80	0.97
3	3.94	5.71
4	0.25	0.30
5	3.94	5.71
6	0.80	0.97
7	3.93	5.69
8	0.14	0.17

size is, for example, 8×8, there is zero gain achieved by temporal filtering when the shifts among the images are integer multiples of 8.

These results suggest that one should not forget that the quantization noise model is merely what its name suggests—a model. True quantization error is not a random process, but rather a deterministic and repeatable quantity; compressing the same image at two different times produces quantization errors that are identical for the two images. However, as discussed in Section 2 the quantization noise model can provide a foundation upon which restoration algorithms can be built, and although the drastic results of Table 5 demonstrate a limitation of the model, the pathological conditions of this subsection's experiment (namely, that all frames differ in global translation by a constant integer shift) rarely occur in natural video. Subsection 5.4 presents results for actual video that demonstrate the quantization noise model's utility in temporal filtering.

5.3. Experiment 3

Here, we repeat the experiment of Subsection 5.1 by replacing the two-dimensional DWT by the three-dimensional DWT. Such a situation was discussed in Section 3, where it was demonstrated that pixel errors' statistical behavior varies both spatially and temporally. Here, three noise models will be compared, all of which are Gaussian: Using the full $\mathbf{K}_{\mathbf{e}_\mathbf{z}^l}$; using an IID model; and using independent noise with variances that are constant within

TABLE 6

Restoration results for compression that uses a three-dimensional DWT. The PSNR listed under "Quantized" is the highest of the PSNR's for the 16 input images, and corresponds to the PSNR at image 8 (compare with image 8 of Figure 2). The noise model "Gaussian Ind" is an independent Gaussian noise model that assumes each received frame has a constant error variance as predicted in Figure 2.

Sequence name	PSNR, dB			
	Quantized	Gaussian IID	Gaussian Ind	Gaussian $\mathbf{K}_{\mathbf{e}_{\mathbf{z}}^{l}}$
barb	36.27	40.19	40.29	42.38
boat	34.01	36.69	36.79	39.17
mandrill	31.30	35.34	35.37	38.50
peppers	35.35	36.77	36.79	37.90
test-pattern	36.24	39.15	39.21	41.14

a frame, but vary between frames in a manner predicted by the plot in Figure 2. As was the case in Figure 2, the length of the transform in the temporal direction is sixteen; the image sequence is synthesized by shifts applied to a prototype image of $(0, 0)$, $(1, 0)$, $(1, 1)$, and continuing in an outward counter-clockwise spiral to the final shift of $(-1, 2)$. All of the sixteen frames in this group of pictures are filtered to produce the estimate of the original image.

Quantitative results for the experiment are shown in Table 6. Significant PSNR improvements are evident relative to the received noisy images, but this is to be expected since the filtering is able to make use of sixteen noisy versions of the exact same original frame; in real-life situations, rarely will such fortuitous circumstances arise. The important conclusions to be drawn from the results are not in regard to PSNR improvement relative to the noisy images, but rather the PSNR improvements of the noise models relative to those of the IID Gaussian noise model. One important result is that relative to the IID noise model's results, there is practically nothing to be gained in temporal filtering by making use of the frame-to-frame error variances that were demonstrated in Figure 2; it seems that the pixel-wise noise characteristics are considerably more important than the frame-wise noise characteristics. The second important result to be drawn from the table is the significant gain of using the full theoretic covariance matrix $\mathbf{K}_{\mathbf{e}_{\mathbf{z}}^{l}}$ relative to the other noise models.

5.4. Experiment with Real Video

While the previous three subsections dealt with synthetically-generated video, such situations are not necessarily indicative of the results one would obtain with real video sequences. This subsection considers temporal filtering of actual motion imagery that has been compressed using the 2D DWT. We will consider two cases: When the observation error is dominated by quantization noise, and hence the motion-compensation error $\mathbf{e}^{k,l}$ can be neglected; and when $\mathbf{e}^{k,l}$ is included in the formulation. For each of the two cases, two models are used for the quantization error—IID Gaussian noise with covariance matrix of $\sigma^2\mathbf{I}$, and Gaussian noise with covariance $\mathbf{K}_{\mathbf{e}_{\mathbf{z}}^l}$.

In previous subsections, the motion between input frames was known exactly because of the artificial construction of the sequences. With real video, the motion must be estimated. Here, we consider video sequences that contain frames that differ by a global transformation, e.g., stationary scenes captured by a moving camera that is located at a far distance from the actual scene, as might be expected from aerial surveillance video. Stationary scenes are not a requirement, but they make motion estimation simpler; with more sophisticated motion estimation, the algorithm described here could be applied equally well. To register frames of the input imagery, an affine motion model is employed. The six parameters of the affine model that relate the two frames are estimated iteratively within a coarse-to-fine multiresolution pyramid. Note that while previous sections assumed integer pixel motions, the more general model here allows for floating-point pixel motions; $\mathbf{A}_{k,l}$ matrices are constructed based on bilinear interpolation for these fractional pixel motions.

Figure 16 shows five frames of the *stickers* sequence; the author acquired this uncompressed sequence using a Pixelink PL-A641 monochrome camera. While the sequence is certainly not the same as aerial surveillance video, it does share certain qualities—a large global-motion component, as well as significant detail at fine resolutions. The writing on the stickers will serve as a sort of resolution chart for comparison of the restoration algorithms. For this example, the compressed versions of these five frames will be used to reconstruct image 2.

Results are first presented for two compression qualities under the assumption that the

Fig. 16. Original 640 × 480 input images of the *stickers* sequence, numbered from image 0 to image 4.

quantization error dominates the overall error $\mathbf{n}^{k,l}$. As in previous subsections, the compression is simulated by scalar quantization of the coefficients from a four-level DWT decomposition. Figure 17 presents results for images compressed to a relatively low quality of approximately 32.3 dB, while Figure 18 presents results for images compressed at higher quality of about 36.5 dB. Each figure contains the compressed middle frame of the five input images, along with restorations using both the full quantization noise model and the IID noise model.

In Figure 17, both the quantization noise model and the IID noise model show considerable improvement over the original compressed image, which shows the potential advantages of temporal filtering for this type of compressed sequence. The visual improvement of the quantization noise model relative to the IID noise model is evidenced by an overall sharper reconstruction, with more apparent contrast. The PSNR is also slightly higher for the restoration using the quantization noise model.

For the higher quality compressed image set of Figure 18, visual distinction among the images is not as clear. There are some improvements, both in visual quality and PSNR, for the restorations compared to the compressed image; for example, the "California" sticker, or the "Go Bananas" sticker. Although difficult to discern on the printed page, the restoration for the quantization noise model is slightly sharper than that of the IID noise model; the "Go Bananas" sticker is arguably more legible for the restoration of the quantization noise model. Unfortunately, the additional sharpness for this higher-quality case comes at the cost of a sharpening of noise as well, which accounts for the slightly lower PSNR for the restoration of the quantization noise model. In this case, noise being sharpened is not compression noise $\mathbf{e}_{\mathbf{z}}^{l}$, but rather the motion compensation noise $\mathbf{e}^{k,l}$. Recall that the restorations of this

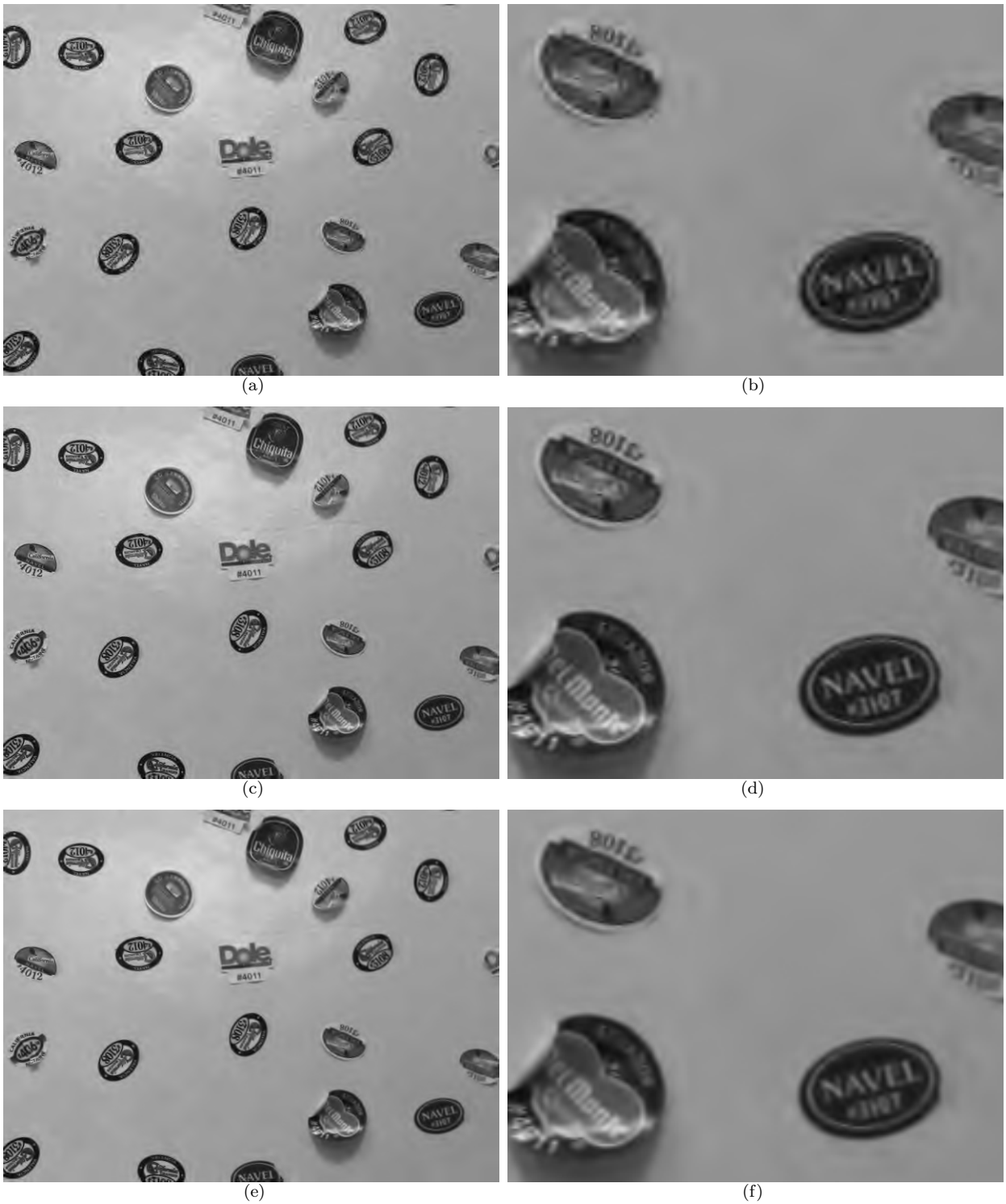


Fig. 17. Restoration results for real video data. The five input images have PSNR values in the range of 32.2 dB to 32.4 dB. The left column shows the full 640 × 480 images, while the right column shows zoomed portions of the images. (a,b): compressed image 2, PSNR=32.23 dB; (c,d): restored image 2 using quantization noise model, PSNR=34.91 dB; (e,f): restored image 2 using IID noise model, PSNR=34.58 dB.

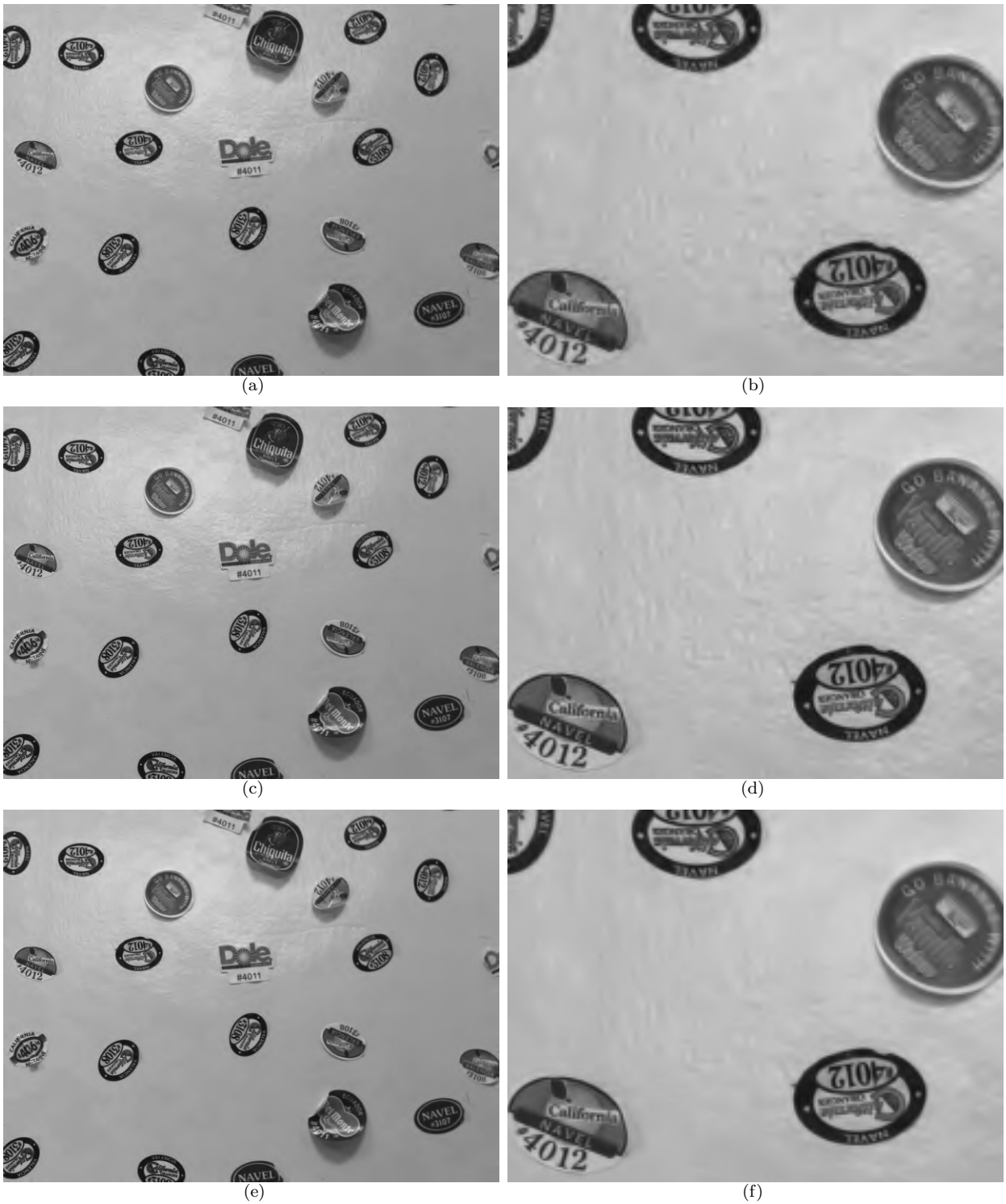


Fig. 18. Restoration results for real video data. The five input images have PSNR values in the range of 36.4 dB to 36.6 dB. The left column shows the full 640 × 480 images, while the right column shows zoomed portions of the images. (a,b): compressed image 2, PSNR=36.44 dB; (c,d): restored image 2 using quantization noise model, PSNR=37.41 dB; (e,f): restored image 2 using IID noise model, PSNR=37.81 dB.

section thus far have not included the motion compensation noise, but rather have assumed that the quantization noise dominates. Since the quantization noise covariance matrix is non-diagonal, and hence includes correlations among the errors, an overall sharpening can result from the filtering. For high-quality compressed motion imagery, the quantization error ceases to dominate as it did in the lower-quality case, and the motion-compensating error becomes more influential. As a result, the quantization noise model sharpens this motion-compensation noise. Such sharpening is not present for the IID case, which assumes that the noise terms at each pixel observation are independent and identically distributed. Thus the very quality that prevents the IID quantization noise model from providing sharp restorations also prevents it from enhancing noise.

To better account for the motion compensation error, one can explicitly model the term $\mathbf{e}^{k,l}$ such that the overall noise term would be

$$\mathbf{n}^{k,l} = \mathbf{e}_{\mathbf{z}}^{l} + \mathbf{e}^{k,l}. \tag{29}$$

The motion compensation noise term $\mathbf{e}^{k,l}$ has often been modeled as IID-Gaussian distributed [22] with variance terms $\lambda^{k,l}$; note that $\lambda^{k,l}$ is zero for $k = l$. With such an assumption, the overall noise term for the IID quantization noise case would have covariance matrix

$$\begin{aligned} \mathbf{K}_{k,l} &= \sigma^2\mathbf{I} + \lambda^{k,l}\mathbf{I} && (30) \\ &= \widehat{\lambda}^{k,l}\mathbf{I}, && (31) \end{aligned}$$

which for practical purposes is nearly equivalent to the IID quantization noise model used previously; this explains why the IID quantization noise model does not enhance the motion compensation error. For the non-IID quantization noise model, the covariance matrix becomes

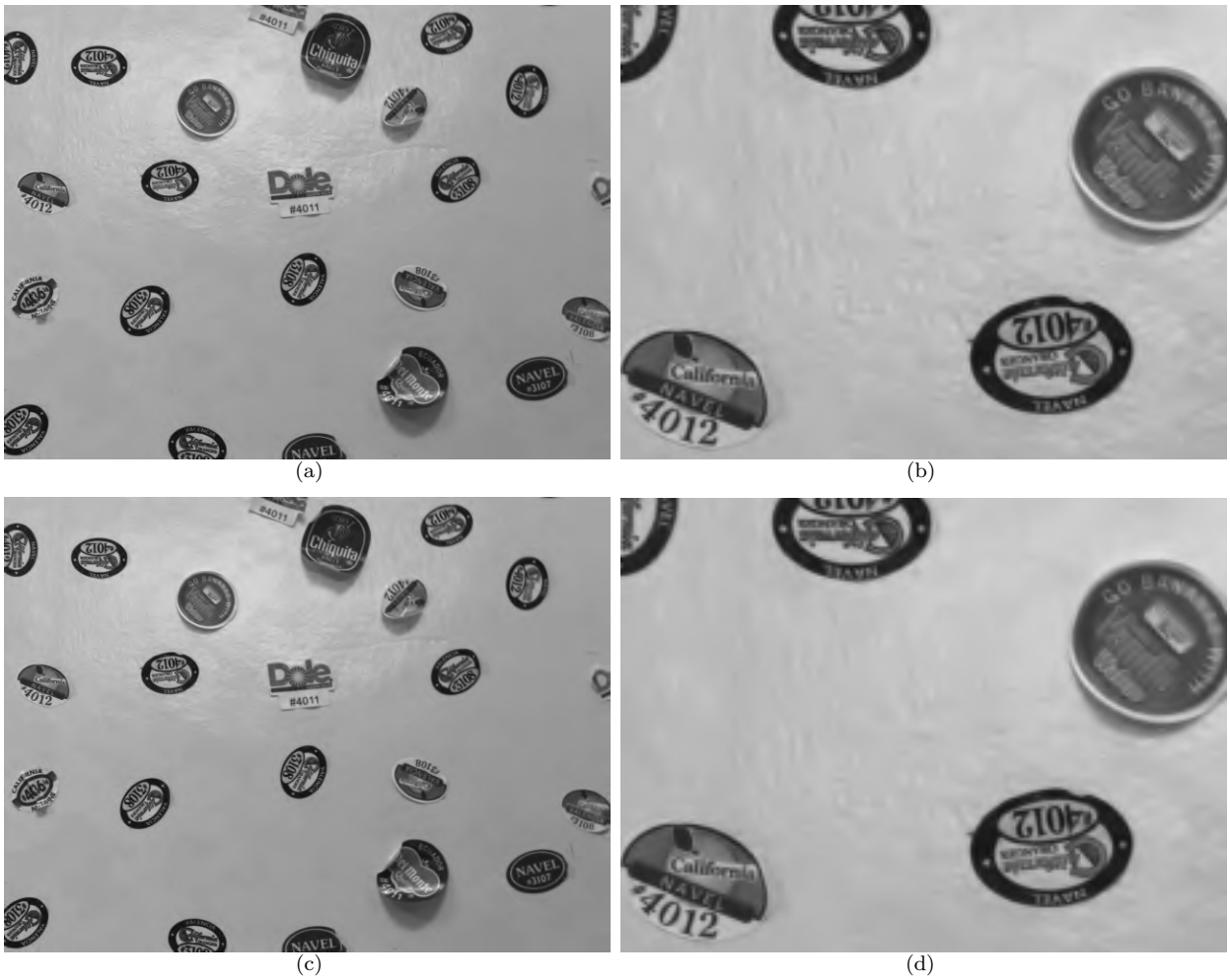
$$\mathbf{K}_{k,l} = \mathbf{H}^{-1}\mathbf{K}_{\mathbf{e}_{\mathbf{y}}^{l}}\mathbf{H}^{-t} + \lambda^{k,l}\mathbf{I}. \tag{32}$$

The restoration algorithms require the inversion of the noise covariance matrix, and it should be readily evident that the inversion for the diagonal IID case is much easier than for the non-diagonal and non-IID case. However, since only the product of the matrix inverse with some input vector is needed, and not the actual explicit inverse matrix, iterative methods

can be used in the non-IID case. Appendix D discusses the iterative implementation for inversion of the covariance matrix in (32).

Explicitly modeling the motion compensation error $\mathbf{e}^{k,l}$ within the overall noise $\mathbf{n}^{k,l}$ leads to improvements in the restorations of both IID and non-IID quantization noise models. Figure 19 shows results for the experiment first reported in Figure 18, but using the modified error covariance matrices of (30) and (32) instead. As indicated in the figure caption, the PSNR performance of the restoration algorithm that used the full quantization noise covariance matrix has risen above that of the IID quantization noise assumption; in both cases, PSNR's are higher than when neglecting the $\mathbf{e}^{k,l}$ term. (Note that since the IID quantization noise assumption already modeled $\mathbf{e}^{k,l}$ fairly well, only marginal improvements result for the IID case.) Although PSNR has been improved, visual differences between these images and their counterparts presented in Figure 18 are difficult to discern. Similarly, the visual improvement of Figure 19(b) relative to Figure 19(d) is comparable to that between Figure 18(d) and Figure 18(f).

A final observation about temporal filtering concerns the algorithms' sensitivity to errors in motion estimation. Motion *compensation* error, as used here, refers to the difference between an image area and its corresponding area in a reference image. Motion *estimation* error refers to inaccuracies in the registration of the areas of the two images—for example, saying that an object has moved by six pixels when in fact it has only moved by four. For the same reasons that the non-IID quantization noise model leads to noise amplifications for motion-compensation error, the non-IID model also leads to error amplification for motion-estimation errors. Errors in motion estimation lead to large errors in motion compensation, which when combined with the sharpening features of the quantization noise model lead to ringing-like artifacts. The IID quantization noise model does not suffer from such problems, because it only averages pixels rather than sharpening them; motion-estimation errors for the IID case simply result in over-blurring, and sometimes introducing ghosting artifacts, in the restoration. Additionally, when using block-based translational motion-estimation techniques for constructing the motion-compensating $\mathbf{A}_{k,l}$ matrices, ringing artifacts can sometimes result near the block boundaries for the restorations using the quantization noise model. Such phenomena do not necessarily indicate a limitation of the quantization noise



(a) (b)

(c) (d)

Fig. 19. Restoration results for real video data using more complex error covariance matrices of (30) and (32). The five input images have PSNR values in the range of 36.4 dB to 36.6 dB. The left column shows the full 640×480 images, while the right column shows zoomed portions of the images. Compressed image 2 was shown in Figure 18(a,b), with PSNR=36.44 dB, and is not re-displayed here. (a,b): Restored image 2 using quantization noise model, PSNR=38.22 dB; (c,d): restored image 2 using IID noise model, PSNR=37.88 dB. Here, $\lambda^{k,l} = 20|k-l|+10$, and $\sigma^2 = 24.1$.

model, but simply reinforce the necessity of accurately estimating correspondences between the input images. Artifacts can be avoided by accurately estimating and assigning the motion-compensation error covariance terms[3] $\lambda^{k,l}$. However, as motion estimation accuracy

[3]In addition, this term can be modified such that the pixel variances vary from position to position depending on the accuracy of the motion estimation.

decreases (or motion compensation error increases), the $\lambda^{k,l}$ by necessity become large enough to dominate the overall error covariance term $\mathbf{K}_{k,l}$. When $\lambda^{k,l}$ dominate, the resulting restorations are nearly identical to those achieved using the IID model. If accurate motion estimation is simply not possible, then temporal filtering with the IID quantization noise model may be more appropriate due to its robustness to errors in motion estimation, in addition to its simpler implementation. However, with accurate motion estimation the quantization noise model provides better restoration results.

6. Conclusion

Pixel errors introduced by quantization of the image's DWT coefficients are not independent, nor are they identically distributed. Experimental observations and theoretic derivations presented here indicate that the errors at different pixel locations have determinedly different characteristics, and the noise covariance matrix developed in this technical report more accurately describes the quantization errors than does an IID noise assumption. As one example application, this report has presented a novel deblurring algorithm that makes use of the quantization noise model within a Bayesian framework, whose superiority to the same Bayesian formulation with assumed IID noise was demonstrated. A second example application was motion imagery restoration by temporal filtering, where the benefits of using the proposed noise model over an IID noise assumption were demonstrated. Appropriate applications of this second example include motion imagery compressed using a 3D DWT, as well as individually-compressed frames of video, for example, Motion JPEG 2000 [28]. The quantization noise model introduced here can find use in many situations where wavelet-compressed imagery must be processed by algorithms that are sensitive to noise.

References

[1] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," in *Int. Conf. on Image Proc.*, vol. 1, Oct. 2001, pp. 185–188.

[2] J. W. Woods and T. Naveen, "A filter based bit allocation scheme for subband compresion of HDTV," *IEEE Trans. on Image Proc.*, vol. 1, no. 3, pp. 436–440, July 1992.

[3] ISO/IEC 15444-1, "Information technology—JPEG 2000 image coding system—Part 1: Core coding system," 2000.

[4] J. Wei, M. Pickering, M. Frater, J. Boman, and W. Zeng, "Boundary artefact reduction using odd tile length and the low pass first convention (OTLPF)," in *Proc. of SPIE Conf. on Applications of Digital Image Proc.*, vol. 4472, 2001, pp. 282–289.

[5] K. Berkner and E. L. Schwartz, "Removal of tile artifacts using projection onto scaling functions for JPEG 2000," in *Int. Conf. on Image Proc.*, vol. 1, Sep. 2002, pp. 373–376.

[6] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. on Image Proc.*, vol. 6, no. 8, pp. 1164–1175, Aug. 1997.

[7] B. Widrow, I. Kollár, and M.-C. Liu, "Statistical theory of quantization," *IEEE Trans. on Instr. and Meas.*, vol. 45, no. 2, pp. 353–361, Apr. 1996.

[8] H. A. Peterson, A. J. Ahumada, and A. B. Watson, "The visibility of DCT quantization noise," in *J. Marreale, ed., Digest of Technical Papers, Society for Information Display, Playa del Rey, CA*, vol. 24, 1993, pp. 942–945.

[9] G. S. Yovanof and S. Liu, "Statistical analysis of the DCT coefficients and their quantization error," in *Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 3–6 1996, pp. 601–605.

[10] T. Meier, K. N. Ngan, and G. Crebbin, "Reduction of blocking artifacts in image and video coding," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 9, no. 3, pp. 490–500, Apr. 1999.

[11] J. Mateos, C. Ilia, B. Jiménez, R. Molina, and A. K. Katsaggelos, "Reduction of blocking artifacts in block transformed compressed color images," in *Int. Conf. on Image Proc.*, vol. 1, Oct. 1998, pp. 401–405.

[12] J. Yang, H. Choi, and T. Kim, "Noise estimation for blocking artifacts reduction in DCT coded images," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 10, no. 7, pp. 1116–1120, Oct. 2000.

[13] A. K. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1989.

[14] S. Kotz, T. J. Kozubowski, and K. Podgórski, *The Laplace Distribution and Generalizations.* Boston: Birkhäuser, 2001.

[15] M. D. Ernst, "A multivariate generalized Laplace distribution," *Computational Statistics*, vol. 13, pp. 227–232, 1998.

[16] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 10, no. 8, pp. 1374–1387, Dec. 2000.

[17] J. Xu, Z. Xiong, S. Li, and Y. Zhang, "Memory-constrained 3-D wavelet transform for video coding without boundary effects," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 12, no. 9, pp. 812–818, Sep. 2002.

[18] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Proc.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[19] J. Biemond, R. L. Lagendijk, and R. M. Mersereau, "Iterative methods for image deblurring," *Proc. of the IEEE*, vol. 78, no. 5, pp. 856–883, May 1990.

[20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analy. and Mach. Intel.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[21] T. P. O'Rourke and R. L. Stevenson, "Improved image decompression for reduced transform coding artifacts," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 5, no. 6, pp. 490–499, Dec. 1995.

[22] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. on Image Proc.*, vol. 5, no. 6, pp. 996–1011, June 1996.

[23] R. L. Stevenson, B. E. Schmitz, and E. J. Delp, "Discontinuity preserving regularization of inverse visual problems," *IEEE Trans. on Syst., Man and Cybern.*, vol. 24, no. 3, pp. 455–469, Mar. 1994.

[24] M. Choi, Y. Yang, and N. P. Galatsanos, "Regularized multichannel recovery of compressed video," in *Int. Conf. on Image Proc.*, vol. 1, Oct. 26–29 1997, pp. 271–274.

[25] Y. Yang, M. Choi, and N. Galatsanos, "New results on multichannel regularized recovery of compressed video," in *Int. Conf. on Image Proc.*, vol. 1, Oct. 4–7 1998, pp. 391–395.

[26] C.-J. Tsai, P. Karunaratne, N. Galatsanos, and A. Katsaggelos, "A compressed video enhancement algorithm," in *Int. Conf. on Image Proc.*, Oct. 23–27 1999.

[27] M. A. Robertson and R. L. Stevenson, "Restoration of compressed video using temporal information," in *SPIE Visual Comm. and Image Proc.*, vol. 4310, Jan. 24–26 2001, pp. 21–29.

[28] ISO/IEC 15444-3, "Information technology—JPEG 2000 image coding system—Part 3: Motion JPEG 2000," 2002.

[29] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY: Springer-Verlag New York, Inc., 1999.

Appendix A
The DWT Matrix and its Transpose

Since the structure of $\mathbf{H}^t$ depends on $\mathbf{H}$, the structure of the DWT matrix $\mathbf{H}$ is first discussed. Only the one-dimensional case is presented; the two- and three-dimensional versions of $\mathbf{H}$ can be constructed by separable extensions of the one-dimensional case. This appendix assumes the Daubechies 9/7 wavelet transform [18], although the case for other wavelet filters is analogous. The low- and high-pass analysis filter coefficients are shown in Table A.1.

For the even-length signal $z[n]$, $n = 0, \ldots, N-1$, a single-level wavelet decomposition can be applied to yield low-pass and high-pass coefficients,

$$\mathbf{y} = \mathbf{H}\mathbf{z}, \tag{33}$$

where the DWT matrix $\mathbf{H}$ is constructed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{lp}^N \\ \mathbf{H}_{hp}^N \end{bmatrix}, \tag{34}$$

where $\mathbf{H}_{lp}^N$ and $\mathbf{H}_{hp}^N$ are low- and high-pass matrix operators of size $\frac{N}{2} \times N$. The low-pass coefficients are taken from the first $\frac{N}{2}$ elements of $\mathbf{y}$, while the high-pass coefficients are taken from the last $\frac{N}{2}$ elements. (For odd N, the given development can be modified accordingly.)

TABLE A.1

Low- and high-pass DWT analysis coefficients.

low-pass		high-pass	
h_4	0.03782845550		
h_3	-0.02384946501	g_3	0.06453888262
h_2	-0.11062440441	g_2	-0.04068941760
h_1	0.37740285561	g_1	-0.41809227322
h_0	0.85269867900	g_0	0.78848561640
h_1	0.37740285561	g_1	-0.41809227322
h_2	-0.11062440441	g_2	-0.04068941760
h_3	-0.02384946501	g_3	0.06453888262
h_4	0.03782845550		

The low- and high-pass matrices are constructed as

$$\mathbf{H}_{lp}^{N} = \begin{bmatrix}
h_0 & 2h_1 & 2h_2 & 2h_3 & 2h_4 & 0 & \cdots & & & & & & & & 0 \\
h_2 & h_1+h_3 & h_0+h_4 & h_1 & h_2 & h_3 & h_4 & 0 & \cdots & & & & & & \\
h_4 & h_3 & h_2 & h_1 & h_0 & h_1 & h_2 & h_3 & h_4 & 0 & \cdots & & & & \\
0 & 0 & h_4 & h_3 & h_2 & h_1 & h_0 & h_1 & h_2 & h_3 & h_4 & 0 & \cdots & & \\
 & & & & & & & & \ddots & & & & & & \\
 & & & \cdots & 0 & h_4 & h_3 & h_2 & h_1 & h_0 & h_1 & h_2 & h_3 & h_4 & 0 \\
 & & & & \cdots & 0 & h_4 & h_3 & h_2 & h_1 & h_0 & h_1 & h_2+h_4 & h_3 & \\
0 & & & & & & & \cdots & 0 & h_4 & h_3 & h_2+h_4 & h_1+h_3 & h_0+h_2 & h_1
\end{bmatrix} \tag{35}$$

and

$$\mathbf{H}_{hp}^{N} = \begin{bmatrix}
g_1 & g_0+g_2 & g_1+g_3 & g_2 & g_3 & 0 & \cdots & & & & & & & & 0 \\
g_3 & g_2 & g_1 & g_0 & g_1 & g_2 & g_3 & 0 & \cdots & & & & & & \\
0 & 0 & g_3 & g_2 & g_1 & g_0 & g_1 & g_2 & g_3 & 0 & \cdots & & & & \\
 & & & & \ddots & & & & & & & & & & \\
 & & & & \cdots & 0 & g_3 & g_2 & g_1 & g_0 & g_1 & g_2 & g_3 & 0 & \\
 & & & & & & \cdots & 0 & g_3 & g_2 & g_1 & g_0 & g_1+g_3 & g_2 & \\
0 & & & & & & & & \cdots & 0 & 2g_3 & 2g_2 & 2g_1 & g_0 &
\end{bmatrix}. \tag{36}$$

The rows of these matrices are shifted versions of a simple prototype low- or high-pass filter with the exception of near the signal boundaries, where the coefficients arise due to symmetric extension of the input signal. Note that slightly different matrices result when the signal length N is odd, and also when the subsampling that corresponds to the wavelet transform is different than even-lowpass—odd-highpass.

Additional levels of wavelet decomposition are achieved by repetitive filtering on resulting low-pass subbands. For example, a two-level wavelet decomposition has a corresponding matrix of

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{lp}^{N/2} & \mathbf{0}^{N/2} \\ \mathbf{H}_{hp}^{N/2} & \\ \hline \mathbf{0}^{N/2} & \mathbf{I}^{N/2} \end{bmatrix} \begin{bmatrix} \mathbf{H}_{lp}^{N} \\ \mathbf{H}_{hp}^{N} \end{bmatrix}, \tag{37}$$

and a three-level wavelet decomposition matrix appears as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{lp}^{N/4} & \mathbf{0}^{N/4} & \mathbf{0}^{N/2} \\ \mathbf{H}_{hp}^{N/4} & & \\ \hline \mathbf{0}^{N/4} & \mathbf{I}^{N/4} & \\ \hline \mathbf{0}^{N/2} & & \mathbf{I}^{N/2} \end{bmatrix} \begin{bmatrix} \mathbf{H}_{lp}^{N/2} & \mathbf{0}^{N/2} \\ \mathbf{H}_{hp}^{N/2} & \\ \hline \mathbf{0}^{N/2} & \mathbf{I}^{N/2} \end{bmatrix} \begin{bmatrix} \mathbf{H}_{lp}^{N} \\ \mathbf{H}_{hp}^{N} \end{bmatrix}, \tag{38}$$

where $\mathbf{I}^N$ and $\mathbf{0}^N$ are the identity and zero matrices of size N. One could continue like this indefinitely, but a three-level example is sufficient for our needs here. When written in the form above, the transpose of the DWT operation is simple to perform; for the three-level decomposition of (38), it is

$$\mathbf{H}^t = \begin{bmatrix} \mathbf{H}_{lp}^{N\,t} & \mathbf{H}_{hp}^{N\,t} \end{bmatrix} \left[\begin{array}{cc:c} \mathbf{H}_{lp}^{N/2\,t} & \mathbf{H}_{hp}^{N/2\,t} & \mathbf{0}^{N/2} \\ \hdashline \multicolumn{2}{c:}{\mathbf{0}^{N/2}} & \mathbf{I}^{N/2} \end{array}\right] \left[\begin{array}{cc:c:c} \mathbf{H}_{lp}^{N/4\,t} & \mathbf{H}_{hp}^{N/4\,t} & \mathbf{0}^{N/4} & \mathbf{0}^{N/2} \\ \hdashline \multicolumn{2}{c:}{\mathbf{0}^{N/4}} & \mathbf{I}^{N/4} & \\ \hdashline \multicolumn{3}{c:}{\mathbf{0}^{N/2}} & \mathbf{I}^{N/2} \end{array}\right]. \quad (39)$$

Each of the three matrices on the right-hand side of (39) consists of the transposes of the low- and high-pass matrices of (35) and (36). Taken individually, the components of $\mathbf{H}^t$ start at the lowest-resolution subbands, apply the individual transpose operation at that subband level, and then apply the transpose operation at the next highest subband level, which continues until all subbands have been processed. One only needs to implement the transposes of (35) and (36), which is readily achieved due to the matrices' sparse and simple structure.

It should be noted that some wavelet transforms are orthogonal, in which case $\mathbf{H}^t\mathbf{H}$ is diagonal; with proper scaling, such DWT matrices become orthonormal, in which case $\mathbf{H}^t$ is implemented by application of the inverse DWT. Many wavelet transforms, while not strictly orthogonal, are designed to be nearly orthogonal; this may tempt the unwary to the use of the inverse DWT in place of the actual transpose. Nevertheless, use of the true transpose is more justified theoretically, yields better results, and since it can be implemented efficiently there is little computational justification for choosing the inverse over the transpose. Use of the transpose also avoids any scaling issues associated with non-unity gains in the various subbands of the wavelet decomposition, as are present in the wavelet transform recommended by the JPEG 2000 standard [3].

Appendix B

Computation of Deblurring Step Size

The objective function being minimized is given by

$$\mathcal{O}(\mathbf{z}) = \frac{1}{2}(\mathbf{Bz} - \mathbf{z}_q)^t \mathbf{K}_{\mathbf{e_z}}^{-1}(\mathbf{Bz} - \mathbf{z}_q) + \lambda \sum_{c \in \mathcal{C}} \rho_T(\mathbf{d}_c^t \mathbf{z}). \tag{40}$$

Given a direction $\mathbf{g}$, the optimal step size is found as

$$\hat{\alpha} = \arg\min_{\alpha} \mathcal{O}(\mathbf{z} - \alpha\mathbf{g}), \tag{41}$$

a one-dimensional minimization along the line $\mathbf{g}$. Written explicitly, we wish to minimize

$$\mathcal{O}(\mathbf{z} - \alpha\mathbf{g}) = \frac{1}{2}(\mathbf{Bz} - \mathbf{z}_q - \alpha\mathbf{Bg})^t \mathbf{K}_{\mathbf{e_z}}^{-1}(\mathbf{Bz} - \mathbf{z}_q - \alpha\mathbf{Bg}) + \lambda \sum_{c \in \mathcal{C}} \rho_T(\mathbf{d}_c^t \mathbf{z} - \alpha \mathbf{d}_c^t \mathbf{g}). \tag{42}$$

Taking the partial derivative with respect to α and setting it equal to zero results in the equation for the optimal α,

$$-\mathbf{g}^t \mathbf{B}^t \mathbf{H}^t \mathbf{K}_{\mathbf{e_y}}^{-1} \mathbf{H}(\mathbf{Bz} - \mathbf{z}_q - \alpha\mathbf{Bg}) - \lambda \sum_{c \in \mathcal{C}} \mathbf{g}^t \mathbf{d}_c \rho_T'(\mathbf{d}_c^t \mathbf{z} - \alpha \mathbf{d}_c^t \mathbf{g}) = 0, \tag{43}$$

where $\mathbf{K}_{\mathbf{e_z}}$ has been expanded according to (3). The term above that corresponds to the prior probability is non-linear, which makes an exact solution difficult. However, by neglecting the higher-order terms of a Taylor representation, the approximation

$$\rho_T'(\mathbf{d}_c^t \mathbf{z} - \alpha \mathbf{d}_c^t \mathbf{g}) \approx \rho_T'(\mathbf{d}_c^t \mathbf{z}) - \alpha \rho_T''(\mathbf{d}_c^t \mathbf{z}) \mathbf{d}_c^t \mathbf{g} \tag{44}$$

arises. Substituting (44) into (43),

$$\begin{aligned} -\mathbf{g}^t \mathbf{B}^t \mathbf{H}^t \mathbf{K}_{\mathbf{e_z}}^{-1} \mathbf{H}(\mathbf{Bz} - \mathbf{z}_q) - \mathbf{g}^t \lambda \sum_{c \in \mathcal{C}} \mathbf{d}_c \rho_T'(\mathbf{d}_c^t \mathbf{z}) + \\ \alpha \mathbf{g}^t \mathbf{B}^t \mathbf{H}^t \mathbf{K}_{\mathbf{e_y}}^{-1} \mathbf{HBg} + \alpha\lambda \mathbf{g}^t \sum_{c \in \mathcal{C}} \mathbf{d}_c \rho_T''(\mathbf{d}_c^t \mathbf{z}) \mathbf{d}_c^t \mathbf{g} = 0, \end{aligned} \tag{45}$$

which simplifies to

$$-\mathbf{g}^t \nabla \mathcal{O}(\mathbf{z}) + \alpha \mathbf{g}^t \left[\mathbf{B}^t \mathbf{H}^t \mathbf{K}_{\mathbf{e_y}}^{-1} \mathbf{HB} + \lambda \sum_{c \in \mathcal{C}} \mathbf{d}_c \rho_T''(\mathbf{d}_c^t \mathbf{z}) \mathbf{d}_c^t \right] \mathbf{g} = 0, \tag{46}$$

where $\nabla \mathcal{O}(\mathbf{z})$ is the gradient of (40) with respect to $\mathbf{z}$, equivalent to the equation given in (19). The final estimate for the optimal step size is then

$$\alpha = \frac{\mathbf{g}^t \nabla \mathcal{O}(\mathbf{z})}{\mathbf{g}^t \left[\mathbf{B}^t \mathbf{H}^t \mathbf{K}_{\mathbf{e_y}}^{-1} \mathbf{HB} + \lambda \sum_{c \in \mathcal{C}} \mathbf{d}_c \rho_T''(\mathbf{d}_c^t \mathbf{z}) \mathbf{d}_c^t \right] \mathbf{g}}. \tag{47}$$

Note that when the direction $\mathbf{g}$ is chosen as the gradient of the objective function, then $\mathbf{g} = \nabla \mathcal{O}(\mathbf{z})$, and the numerator of the above equation matches that given earlier in (20).

Appendix C
Optimization for Temporal Filtering

To optimize (27), the method of conjugate gradients is used. For iteration $i+1$, a new estimate for $\mathbf{z}^k$ is obtained from the estimate at iteration i as

$$\mathbf{z}^k_{(i+1)} = \mathbf{z}^k_{(i)} - \alpha_{(i)}\mathbf{d}_{(i)}, \tag{48}$$

where $\mathbf{z}^k_{(i)}$ is the estimate at iteration i, $\mathbf{d}_{(i)}$ is a direction vector and $\alpha_{(i)}$ is the step size that controls how far along $\mathbf{d}_{(i)}$ from $\mathbf{z}^k_{(i)}$ the new estimate is taken.

The direction is computed as

$$\mathbf{d}_{(i)} = \mathbf{g}_{(i)} + \beta_{(i)}\mathbf{d}_{(i-1)}, \tag{49}$$

where $\beta_{(i)}$ is

$$\beta_{(i)} = \frac{\mathbf{g}^t_{(i)}\mathbf{g}_{(i)}}{\mathbf{g}^t_{(i-1)}\mathbf{g}_{(i-1)}}, \tag{50}$$

and $\mathbf{g}_{(i)}$ is the gradient of the objective function in (27), computed as

$$\mathbf{g}_{(i)} = \sum_{l=k-n}^{k+n} \mathbf{A}^t_{k,l}\mathbf{K}^{-t}_{k,l}(\mathbf{A}_{k,l}\mathbf{z}^k_{(i)} + \boldsymbol{\mu}^{k,l} - \mathbf{z}^l_q). \tag{51}$$

In the simplifying case where $\mathbf{K}_{k,l} = \mathbf{K}_{\mathbf{e}^l_{\mathbf{z}}} = \mathbf{H}^{-1}\mathbf{K}_{\mathbf{e}^l_{\mathbf{y}}}\mathbf{H}^{-t}$, the gradient becomes

$$\mathbf{g}_{(i)} = \sum_{l=k-n}^{k+n} \mathbf{A}^t_{k,l}\mathbf{H}^t\mathbf{K}^{-1}_{\mathbf{e}^l_{\mathbf{y}}}\mathbf{H}(\mathbf{A}_{k,l}\mathbf{z}^k_{(i)} + \boldsymbol{\mu}^{k,l} - \mathbf{z}^l_q). \tag{52}$$

The initial estimate for the image is taken as $\mathbf{z}^k_{(0)} = \mathbf{z}^k_q$, and the initial direction is taken as $\mathbf{d}_{(0)} = \mathbf{g}_{(0)}$.

The optimal step size $\alpha_{(i)}$ is computed in a manner similar to that of the deblurring case, which was presented in Appendix B. The result of a comparable derivation is

$$\alpha_{(i)} = \frac{\mathbf{d}^t_{(i)}\mathbf{g}_{(i)}}{\mathbf{d}^t_{(i)}\sum_{l=k-n}^{k+n} \mathbf{A}^t_{k,l}\mathbf{K}^{-1}_{k,l}\mathbf{A}_{k,l}\mathbf{d}_{(i)}}, \tag{53}$$

which, for the case $\mathbf{K}_{k,l} = \mathbf{K}_{\mathbf{e}^l_{\mathbf{z}}} = \mathbf{H}^{-1}\mathbf{K}_{\mathbf{e}^l_{\mathbf{y}}}\mathbf{H}^{-t}$, becomes

$$\alpha_{(i)} = \frac{\mathbf{d}^t_{(i)}\mathbf{g}_{(i)}}{\mathbf{d}^t_{(i)}\sum_{l=k-n}^{k+n} \mathbf{A}^t_{k,l}\mathbf{H}^t\mathbf{K}^{-1}_{\mathbf{e}^l_{\mathbf{y}}}\mathbf{H}\mathbf{A}_{k,l}\mathbf{d}_{(i)}}. \tag{54}$$

For more details on the conjugate gradient method of optimization, consult [29].

Appendix D
Inversion of Full Observation Error Covariance Matrix

The inverse of the full observation error covariance matrix $\mathbf{K}_{k,l}$ is needed to implement the optimization of Appendix C when $\lambda^{k,l} \neq 0$,

$$\mathbf{K}_{k,l}^{-1} = \left[\mathbf{H}^{-1}\mathbf{K}_{\mathbf{e}_\mathbf{y}^l}\mathbf{H}^{-t} + \lambda^{k,l}\mathbf{I}\right]^{-1}. \tag{55}$$

Due to the combination of the quantization noise covariance matrix and the motion compensation error covariance matrix, the above inverse cannot be implemented through simple operations as is possible when only one of the two noise terms is present. Since the inverse matrix is always applied to an input vector $\mathbf{v}$ to yield an output vector $\mathbf{u}$, it suffices to solve only for the vector $\mathbf{u}$ without explicitly evaluating and storing the full inverse matrix,

$$\mathbf{u} = \left[\mathbf{H}^{-1}\mathbf{K}_{\mathbf{e}_\mathbf{y}^l}\mathbf{H}^{-t} + \lambda^{k,l}\mathbf{I}\right]^{-1}\mathbf{v}, \tag{56}$$

which is equivalent to solving the following equation for $\mathbf{u}$:

$$\left[\mathbf{H}^{-1}\mathbf{K}_{\mathbf{e}_\mathbf{y}^l}\mathbf{H}^{-t} + \lambda^{k,l}\mathbf{I}\right]\mathbf{u} = \mathbf{v}. \tag{57}$$

Equation (57) is solved here using the conjugate gradient method. For notational convenience, replace the matrix to be inverted by the matrix $\mathbf{P}$, such that (57) becomes

$$\mathbf{P}\mathbf{u} = \mathbf{v}. \tag{58}$$

Solving (58) is equivalent to minimizing

$$\frac{1}{2}\mathbf{u}^t\mathbf{P}\mathbf{u} - \mathbf{u}^t\mathbf{v}, \tag{59}$$

whose minimum is found when the gradient $\mathbf{g}$ with respect to $\mathbf{u}$ is zero,

$$\mathbf{g} = \mathbf{P}\mathbf{u} - \mathbf{v} = \mathbf{0}. \tag{60}$$

The $\mathbf{u}$ that minimizes (59) is computed using the method of conjugate gradients in the exact same manner as was described in Appendix C. Equations (49) and (50) are used to determine a direction $\mathbf{d}_{(i)}$, which when used with a step size of

$$\alpha_{(i)} = \frac{\mathbf{d}_{(i)}^t\mathbf{g}_{(i)}}{\mathbf{d}_{(i)}^t\mathbf{P}\mathbf{d}_{(i)}} \tag{61}$$

leads to an updated estimate for $\mathbf{u}$, analogous to the update equation of (48). Note that in application of $\mathbf{P}$, the transpose of the inverse DWT must be evaluated, i.e., $\mathbf{H}^{-t}$. Such a function is implemented similarly to the case for $\mathbf{H}^t$, which was discussed in Appendix A.

$\mathbf{0}$	zero matrix or vector
$\mathbf{0}^N$	zero matrix of size $N \times N$
$\mathbf{A}_{k,l}$	matrix operator for predicting frame $\mathbf{z}^l$ from $\mathbf{z}^k$
$\mathbf{B}$	blurring matrix operator
$\mathbf{C}$	parameter of multivariate Laplace distribution
$\mathcal{C}$	set of all cliques in HMRF
CCD	charge coupled device
DCT	discrete cosine transform
DWT	discrete wavelet transform
$E_Y[u, v]$	$(u, v)^{th}$ element of DWT error represented by $\mathbf{e_y}$
GMRF	Gauss-Markov random field
$\mathbf{H}$	DWT matrix
$\mathbf{H}^t$	transpose of matrix $\mathbf{H}$
$\mathbf{H}^{-1}$	inverse of matrix $\mathbf{H}$
$\mathbf{H}_{hp}^N$	high-pass matrix operator
$\mathbf{H}_{lp}^N$	low-pass matrix operator
HMRF	Huber-Markov random field
$\mathbf{I}$	identity matrix
$\mathbf{I}^N$	identity matrix of size $N \times N$
IID	independent and identically distributed
$\mathbf{K}_{\mathbf{e_y}}$	quantization error covariance matrix for $\mathbf{e_y}$
$\mathbf{K}_{\mathbf{e_z}}$	quantization error covariance matrix for $\mathbf{e_z}$
$\mathbf{K}_{\mathbf{e}_\mathbf{y}^l}$	quantization error covariance matrix for $\mathbf{e}_\mathbf{y}^l$
$\mathbf{K}_{\mathbf{e}_\mathbf{z}^l}$	quantization error covariance matrix for $\mathbf{e}_\mathbf{z}^l$
$\mathbf{K}_{k,l}$	error covariance matrix for observations $\mathbf{z}_q^l$ given $\mathbf{z}^k$
$\mathbf{K_y}$	covariance matrix of DWT decomposition $\mathbf{y}$
$\mathbf{K_z}$	covariance matrix of image $\mathbf{z}$
$K_{\mathbf{e_z}}[m_1, m_2]$	$(m_1, m_2)^{th}$ element of $\mathbf{K}_{\mathbf{e_z}}$

$K_{N,\eta}$	constant term in multivariate Laplace pdf
MAP	maximum a posteriori
ML	maximum likelihood
MSE	mean-squared error
$\mathcal{O}(\cdot)$	objective function being minimized
PSNR	peak signal to noise ratio
PSNR_c	PSNR at convergence in deblurring algorithm
PSNR_m	maximum PSNR achieved by deblurring algorithm
T	Huber parameter
V	normalization term in HMRF
$Y[u,v]$	$(u,v)^{th}$ element of DWT decomposition represented by $\mathbf{y}$
$Z[m,n]$	$(m,n)^{th}$ element of image represented by $\mathbf{z}$
$\mathbf{d}$	direction vector in conjugate gradient optimization
$\mathbf{d}_c^t$	when applied to $\mathbf{z}$, extracts second-order pixel differences for HMRF
$\text{diag}(\cdot)$	vector formed by taking diagonal of the matrix argument
$\mathbf{e}^{k,l}$	error in predicting $\mathbf{z}^l$ from $\mathbf{z}^k$
$\mathbf{e}_\mathbf{y}$	quantization error for $\mathbf{y}$
$\mathbf{e}_\mathbf{y}^l$	quantization error for $\mathbf{y}^l$
$\mathbf{e}_\mathbf{z}$	quantization error for $\mathbf{z}$
$\mathbf{e}_\mathbf{z}^l$	quantization error for $\mathbf{z}^l$
$\mathbf{g}$	gradient of the objective function being minimized
$\mathbf{h}_{u,v}^{-1}$	basis image of $(u,v)^{th}$ DWT coefficient
$h_{u,v}^{-1}[m]$	m^{th} element of $\mathbf{h}_{u,v}^{-1}$
$\left[\mathbf{h}_{u,v}^{-1}\right]^2$	error variance basis images
$\mathbf{n}_a$	acquisition noise
$\mathbf{n}^{k,l}$	error in predicting $\mathbf{z}_q^l$ given $\mathbf{z}^k$
$p(\cdot)$	probability distribution function
pdf	probability distribution function
$\mathbf{y}$	DWT decomposition of $\mathbf{z}$
$\mathbf{y}_q$	DWT coefficients after quantization

$\mathbf{z}$	image vector formed by stacking image columns
$\mathbf{z}_b$	blurred image
$\mathbf{z}_{(i)}$	parenthesized subscript denotes quantities at iteration i of an iterative algorithm
$\mathbf{z}^k$	image vector at frame number k
$\mathbf{z}_q$	image after quantization of its DWT coefficients
Δ_b	quantization step size for b^{th} DWT subband
Δ_d	nominal quantizer
Δ_i	quantization step size for i^{th} DWT coefficient
$\mathbf{\Phi}(\omega)$	characteristic function
α	step size for iterative optimizations
β	parameter in conjugate gradient optimization
γ_b	sum of squared image errors due to unit error in DWT subband b
η	parameter of multivariate Laplace distribution
λ	regularization parameter in HMRF
$\lambda^{k,l}$	variance for IID Gaussian motion-compensation noise
$\boldsymbol{\mu}^{k,l}$	term to account for unobservable pixels when predicting $\mathbf{z}^l$ from $\mathbf{z}^k$
$\rho_T(\cdot)$	Huber penalty function, parameterized by T
$\rho'_T(\cdot)$	first derivative of $\rho_T(\cdot)$
$\rho''_T(\cdot)$	second derivative of $\rho_T(\cdot)$
σ^2	variance for IID Gaussian quantization noise
$\sigma^2_Y[u,v]$	quantization error variance for $(u,v)^{th}$ DWT coefficient